THEORETICAL BIOLOGY AND
MEDICAL MODELLING

RESEARCH   **Open Access**

A symmetry model for genetic coding via a wallpaper group composed of the traditional four bases and an imaginary base E: Towards category theory-like systematization of molecular/genetic biology

Jitsuki Sawamura[1*], Shigeru Morishita[2] and Jun Ishigooka[1]

* Correspondence:
sawamura.jitsuki@twmu.ac.jp
[1]Department of Psychiatry, Tokyo
Women's Medical University,
Tokyo, Japan
Full list of author information is
available at the end of the article

Abstract

Background: Previously, we suggested prototypal models that describe some clinical states based on group postulates. Here, we demonstrate a group/category theory-like model for molecular/genetic biology as an alternative application of our previous model. Specifically, we focus on deoxyribonucleic acid (DNA) base sequences.

Results: We construct a wallpaper pattern based on a five-letter cruciform motif with letters C, A, T, G, and E. Whereas the first four letters represent the standard DNA bases, the fifth is introduced for ease in formulating group operations that reproduce insertions and deletions of DNA base sequences. A basic group $Z_5 = \{r, u, d, l, n\}$ of operations is defined for the wallpaper pattern, with which a sequence of points can be generated corresponding to changes of a base in a DNA sequence by following the orbit of a point of the pattern under operations in group Z_5. Other manipulations of DNA sequence can be treated using a vector-like notation 'D_j' corresponding to a DNA sequence but based on the five-letter base set; also, 'D_j's are expressed graphically. Insertions and deletions of a series of letters 'E' are admitted to assist in describing DNA recombination. Likewise, a vector-like notation R_j can be constructed for sequences of ribonucleic acid (RNA). The wallpaper group $B = \{Z_5^{\times\infty}, \bullet\}$ (an ∞-fold Cartesian product of Z_5) acts on D_j (or R_j) yielding changes to D_j (or R_j) denoted by '$D_j \circ B_{(j \to k)} = D_k$' (or '$R_j \circ B_{(j \to k)} = R_k$'). Based on the operations of this group, two types of groups—a modulo 5 linear group and a rotational group over the Gaussian plane, acting on the five bases—are linked as parts of the wallpaper group for broader applications. As a result, changes, insertions/deletions and DNA (RNA) recombination (partial/total conversion) are described. As an exploratory study, a notation for the canonical "central dogma" via a category theory-like way is presented for future developments.

Conclusions: Despite the large incompleteness of our methodology, there is fertile ground to consider a symmetry model for genetic coding based on our specific wallpaper group. A more integrated formulation containing "central dogma" for future molecular/genetic biology remains to be explored.

Keywords: DNA (RNA) bases, Imaginary base, Wallpaper group, Operation, Cartesian vector, Category, Central dogma

Background

Group theory is the cornerstone in classifying and studying abstract concepts involving symmetry [1,2]. In general, when group theory is used in various fields of natural sciences, it plays an important role in describing geometrical or dynamical symmetries of phenomena under consideration; examples include mathematics [3,4], physics [5-8], chemistry [9], molecular/genetic biology [10-22], and anthropology [23]. Moreover, much fertile ground still exists where group theory can display its versatility from a multitude of viewpoints. To our knowledge, one such candidate is molecular/genetic biology where group theory has already provided great contributions [10-22].

Deoxyribonucleic acid (DNA) is a nucleic acid containing genetic instructions coded in ordered sequences of four bases located in genes that determine specific genetic characteristics of an organism. In the canonical Watson-Crick DNA base pairing, adenine (A) forms a base pair with thymine (T) and guanine (G) forms a base pair with cytosine (C) [24-26]. Similarly, ribonucleic acid (RNA), which has various biological roles, is a molecule that has a much shorter chain of nucleotides. The sequence of DNA consisting of bases 'A, C, T and G' is transcribed into RNA, composed of bases 'A, C, U and G'; the sets differ in that 'U (uracil)' replaces 'T (thymine)'.

Over the latter half of the 20th century, the nature of the genetic code became fairly well established. As for the coding sequences of DNA into nucleotide units, one needs to build up more general, sophisticated, rationally functionalized systematics concerning DNA base sequences that will enable genes to be understood at the molecular biology level in more optimized form. Indeed, many approaches have been undertaken to describe gene characteristics from various viewpoints within the participating disciplines [24-42]. In particular, the concept of 'symmetry' for DNA sequences plays an important role in understanding their characteristics.

However, each has its advantages and disadvantages in terms of utility and convenience in applications. To our knowledge, so far, if we intend to incorporate a sequence of bases into another sequence and/or exclude certain bases from that substitution, we need to look further afield because normally, sequencing and inserting-deleting operations cannot help in distinguishing one from the other. That means that multiple types of operations are necessary if features of DNA containing exceptional sequences are to be treated.
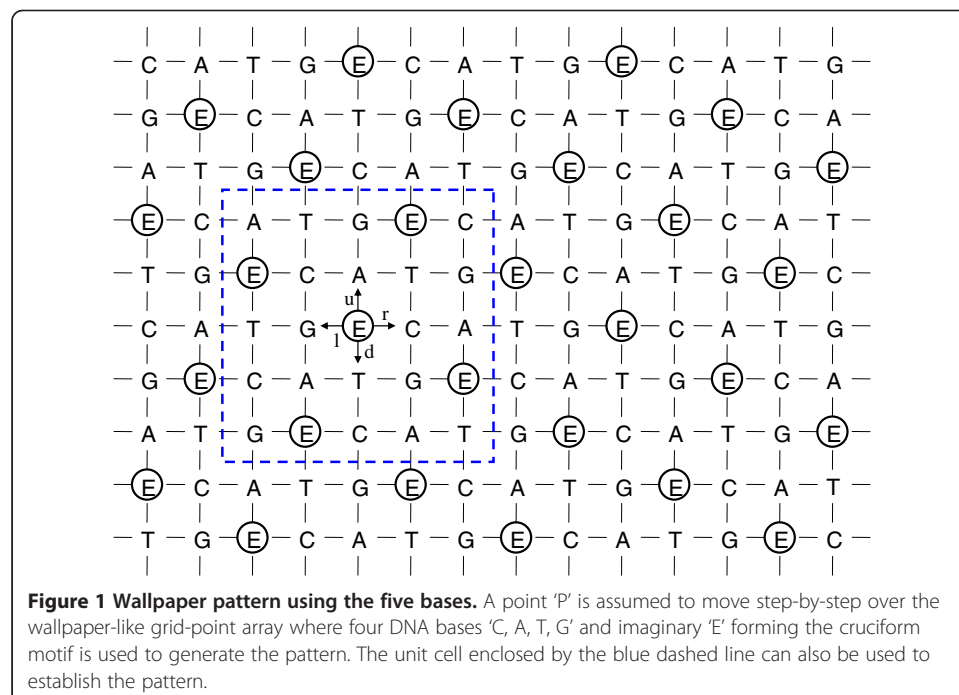
Previously, we suggested prototypal models that describe some clinical states based on group postulates [43]. In this article, we demonstrate a group/category theory-like model for molecular/genetic biology as an alternative application of our previous model. Specifically, focusing on DNA base sequences, we present a simple model where not only changes in sequences of DNA bases but also insertion, deletion, and recombination (partial/total conversion) of DNA bases are treatable within some simple rules via the combination of a set and a group defined over some specific wallpaper pattern. Moreover, a category theory-like formalism, where a description of the DNA bases and their transcription to RNA bases can be made, is attempted from which a category theory-like framework is constructed requiring as few and as simple rules as possible. As an example, by assimilating the canonical "central dogma" [26], we hope to provoke more interactivity among those interested branches of natural science, if possible. The methodology consists of eight parts, the content of which is built-up step-by-step as scope is enlarged to encompass the more advanced themes.

§1 A preliminary setting describing a wallpaper pattern used as a symmetry model for DNA sequences

First, we consider a certain wallpaper pattern that helps us to visualize the operations of the present model (see Figure 1) [2,44-47]. There, the pattern comprises repetitions of a cruciform motif with each motif consisting of five letters E, C, A, T, and G with the latter four letters equally spaced at the points of a cross about a central E. The motif generates the pattern through a translation specified as a knight's move in chess—two steps out and one step right. In this way, the grid-points in this regular wallpaper pattern can be obtained uniquely and be extended indefinitely. Note also that each horizontal line is generated by repetitions of the sequence E-C-A-T-G. Moreover, the line above is a displaced copy of the one below with letter A placed directly above letter E. This preserves the condition that any cruciform is composed of one each of the five letters.

The wallpaper pattern as an array of cruciforms is capable of being constructed as stacks of a unit cell (the 5×5 square enclosed in the dotted line in Figure 1) by horizontal and vertical translations [2,44-47]. The positions of the bases of the cruciform motif are so determined to make it easier to determine the complementary base of each base; the practical applications are clarified later. We introduce the letter 'E' to indicate an 'empty' base which is treated in the same way as the other bases at least for display purposes. This five-base scheme is adopted to aid the notion of group composition in our model.

In addition, we focus on a point 'P' on the wallpaper pattern (i.e., the grid-point array in Figure 1), to compose a certain DNA base sequence. In accordance with this, we shall always adjoin a series of letters that are determined as a trajectory of the point 'P'—also called the 'orbit of P'—over the wallpaper pattern. For instance, when we identify or recognize some changes of DNA bases with 'P' moving from 'A → C → E'

Figure 1 Wallpaper pattern using the five bases. A point 'P' is assumed to move step-by-step over the wallpaper-like grid-point array where four DNA bases 'C, A, T, G' and imaginary 'E' forming the cruciform motif is used to generate the pattern. The unit cell enclosed by the blue dashed line can also be used to establish the pattern.

over the wallpaper pattern, then this represents a series of changes to one base located at a specific position of a DNA sequence in the manner 'ACE...' or '...A...' → '...C...' → '...E...'. The 'orbit of P' can describe series of sequences of DNA bases, or series of changes of each letters in the same places, although, in this article, we focus mainly on the latter case, without provisory context.

With these postulates, we consider the set $C_5 = \{C, A, T, G, E\}$. If the point 'P' moves onto an 'E', 'E' must be included and identified in the series of letters, as in 'ACG**E**T', for example. This is interpreted as the series of DNA bases 'ACGT'. Thus, 'E' depends on context; that is, 'E' can be inserted or removed from any series where we would like to include or eliminate 'E's so long as these are recognized/tracked in the entire process. When read from left to right, the place number of each letter in the series is subscripted, as in '$A_1C_2G_3T_4$'. After insertions/deletions, the place number is augmented/diminished depending on initial and final positions; hence following three insertions '$A_1C_2G_3T_4$' → '$A_1\mathbf{E_2}C_3G_4\mathbf{E_5E_6}T_7$'; this means the point 'P' takes the place 'E' once between A_1 and C_3, and twice between G_4 and T_7 over the wallpaper pattern in Figure 1. More details are to be given later.

As a further refinement, the orbit of 'P' can be stated as a sequence of shift operations as follows; let 'r' denote a move one step to the right corresponding to say A → T, T → G or G → E. Similarly, we denote 'l': move one step to the left as for C → E, and E → G; 'u': move one step up; and 'd': move down. We include 'n' to designate a 'no move' (remain at the same point). A sequence of 'r', 'u', 'd', 'l', and 'n' then provides a position-independent means to describe the orbit of 'P'; any of these five operations can be applied to any of the five letters. We denote their operations on 'P' in the following way. If point 'P' moves from 'E' to 'C' (step to the right), we write 'E∘r = C' where '∘' signifies apply 'r' to 'E' (see Figure 1). In a similar way, 'E∘l = G', 'E∘u = A', 'E∘d = T' and 'E∘n = E'. Note though that each operator means a change of one base to another base within these five bases; the meaning of '=' is not the degree of translation but equivalence to the resultant base from the wallpaper pattern.

To shorten multiple applications of the operations, we introduce '●' to denote the composition of two operations, for example, '((E∘r)∘u) = E∘r●u'. From Figure 1, we find 'E∘d = T' yields the same change as 'E∘r●u = T'. As other examples, 'r●r●d = n' results in 'r●r = u', and 'd●d●l = n' results in 'd●d = r', because from Figure 1, 'E∘r●r = E∘u = A', and 'E∘d●d = E∘r = C'. All possible one-step changes between letters 'C, A, T, G, E' and operators 'r, u, d, l, n', and all possible compositions of operators for the wallpaper pattern of Figure 1 are presented in Figures 2 and 3, and Appendix A.

The binary compositions among the five operations 'r, u, d, l and n' can be shown to satisfy the Abelian group postulates (wallpaper group/plane symmetry group/plane crystallographic group [2-4,44-47]). Indeed, let $Z_5 = \{r, u, d, l, n\}$, then $\{Z_5, ●\}$ is the Abelian group of order five. That is, for all elements $\in Z_5$, we have:

1) Associativity: x●(y●z) = (x●y)●z, (x, y or z being arbitrary elements belonging to Z_5);
2) Identity: 'n' is an identity element such that x●n = n●x = x;
3) Inverse: a unique element x^{-1} exists such that x● x^{-1} = x^{-1}●x = n (x^{-1} is called the inverse element of x);
4) Commutativity: x●y = y●x,
5) Closure: any combination of operations between x●y belongs to Z_5.

$\circ$	r $(\leftrightarrow\omega_1)$	u $(\leftrightarrow\omega_2)$	d $(\leftrightarrow\omega_3)$	l $(\leftrightarrow\omega_4)$	n $(\leftrightarrow\omega_5=\omega_0)$
E	C	A	T	G	E
C	A	T	G	E	C
A	T	G	E	C	A
T	G	E	C	A	T
G	E	C	A	T	G

'↔' : bijection

Figure 2 Cayley tables for the five bases and five operations of linear/rotational groups. Any of the five operations on any of five bases yields a base in a cyclic order.

$\bullet$	r $(\leftrightarrow\omega_1)$	u $(\leftrightarrow\omega_2)$	d $(\leftrightarrow\omega_3)$	l $(\leftrightarrow\omega_4)$	n $(\leftrightarrow\omega_5=\omega_0)$
r $(\leftrightarrow\omega_1)$	u $(\leftrightarrow\omega_2)$	d $(\leftrightarrow\omega_3)$	l $(\leftrightarrow\omega_4)$	n $(\leftrightarrow\omega_5=\omega_0)$	r $(\leftrightarrow\omega_1)$
u $(\leftrightarrow\omega_2)$	d $(\leftrightarrow\omega_3)$	l $(\leftrightarrow\omega_4)$	n $(\leftrightarrow\omega_5=\omega_0)$	r $(\leftrightarrow\omega_1)$	u $(\leftrightarrow\omega_2)$
d $(\leftrightarrow\omega_3)$	l $(\leftrightarrow\omega_4)$	n $(\leftrightarrow\omega_5=\omega_0)$	r $(\leftrightarrow\omega_1)$	u $(\leftrightarrow\omega_2)$	d $(\leftrightarrow\omega_3)$
l $(\leftrightarrow\omega_4)$	n $(\leftrightarrow\omega_5=\omega_0)$	r $(\leftrightarrow\omega_1)$	u $(\leftrightarrow\omega_2)$	d $(\leftrightarrow\omega_3)$	l $(\leftrightarrow\omega_4)$
n $(\leftrightarrow\omega_5=\omega_0)$	r $(\leftrightarrow\omega_1)$	u $(\leftrightarrow\omega_2)$	d $(\leftrightarrow\omega_3)$	l $(\leftrightarrow\omega_4)$	n $(\leftrightarrow\omega_5=\omega_0)$

'↔' : bijection

Figure 3 Cayley tables for the linear group, rotational group and wallpaper group for five bases. This confirms the bijection between the wallpaper group and rotational group.

Therefore, Z_5 is an Abelian group [2-4,44-47]. The inverses for each of the elements are:

$$r^{-1} = l, l^{-1} = r, u^{-1} = d, d^{-1} = u, n^{-1} = n, \tag{1}$$

which can be used to complete the composition table—also known as the Cayley table of the group.

We further stipulate that when we perform these operations, then we always assume/identify the coding of the sequence of DNA bases in accordance with these operations, and vice versa. This is because the action of 'u' on E yields base 'A', that of 'd' on 'E' yields base 'T', that of 'l' on 'E' yield base 'G', and naturally that of 'n' on 'E' results in the same 'E'. For a more complex example, we might insert a certain series of 'A's in 'ACCGT' between the 3rd and 4th base. To begin, we decide to write this manipulation as follows: 'ACC( )GT' is transformed into 'ACC(**E**)GT' by inserting 'E'. Next, because the operation 'u' to the new 4th component 'E' yields 'E → A' ('ACCGT' → 'ACC**A**GT'), and vice versa, that is, 'd' operating on the 4th component 'A' produces 'A → E' ('ACC**A**GT' → 'ACCGT'). In this way, appropriate use of 'E's through the adequate combination of operators of Z_5 enables to express inclusion and/or exclusion of any base between bases in a DNA sequence. To indicate this, we adopt a vector-like description with an infinite number of 'E's being assumed to be present at the end of any given base sequence. This means the point 'P' takes 'E's an infinite number of times over the wallpaper pattern (Figure 1); i.e.,

$$
\begin{aligned}
D_j &= [C|T|G|A|T|A|A|C|\mathbf{E}|\mathbf{E}|\mathbf{E}|\mathbf{E}|\mathbf{E}|...] \\
&= [C_1|T_2(\ )G_3|A_4|T_5(\ )A_6|A_7|C_8|\mathbf{E}_9|\mathbf{E}_{10}|\mathbf{E}_{11}|\mathbf{E}_{12}|\mathbf{E}_{13}|\mathbf{E}_{14}|...]
\end{aligned}
\tag{2a}
$$

$$= [C_1|T_2(\mathbf{E}_3)G_4|A_5|T_6(\mathbf{E}_7|\mathbf{E}_8)A_9|A_{10}|C_{11}|\mathbf{E}_{12}|\mathbf{E}_{13}|\mathbf{E}_{14}|\mathbf{E}_{15}|\mathbf{E}_{16}|...]. \tag{2b}$$

(j: the number of the sequence, N: the number of single-stranded DNA bases of 'D_j's except for the infinite tail of 'E's; in the above case, N = 8)

In the last expression (2b), 'E_3' is inserted before the 3rd component 'G_3' and 6th component 'A_6' marked by '( )' in formula (2a), and the place numbers of all components to the right of the 3rd component are all incremented by '1', those to the right of the 6th component; by '1 + 2'. Likewise, we assume that the deletion of any 'E's that are already displayed in D_j is always permissible according to need with the place numbers being decreased by the necessary size.

Essentially, we regard the subscripted place number of a component of D_j, e.g., '**3**' of 'A**3**', as a convenient place mark to help in recognizing and counting the order of sequences. Place numbers remain fixed when performing operations within a series of operations during code recognition of bases. However, for an operation, another place number is always permissible in principle, from where indexing of a specific DNA base sequence starts.

Alternatively, we use the following notation to describe various cases:

1) we denote by '{D_j}' a sequence 'D_j' where specified 'E's other than the trailing series of 'E's are implicitly implied but the place number indexing is retained; i.e.,

$$\{D_j\} = [C_1|T_2|G_4|A_5|T_6|A_9|A_{10}|C_{11}|E_{12}|E_{13}|E_{14}|E_{15}|E_{16}|...].$$

$$\tag{3a}$$

Here, the explicitly indicated place numbers are the same as in (2b) and missing subscripted place numbers indicate omitted 'E's. Hence, (3a) without trailing 'E's and subscripts represents an ordinal/conventional DNA sequence.

2) we denote by '$<D_j>$' a sequence 'D_j' where specified explicit 'E's other than the trailing 'E's are deleted (changed into implicit 'E's) and the base sequence is re-indexed with sequential place numbers, i.e.,

$$< D_j >= [C_1|T_2|G_3|A_4|T_5|A_6|A_7|C_8|E_9|E_{10}|E_{11}|E_{12}|E_{13}|...]. \tag{3b}$$

Note that 'E's other than the trailing 'E's are not recognized as explicit components and hence are not indexed. Additional insertions/deletions of 'E's are permitted after deletions of 'E's; therefore, apart from the trailing 'E's, (3b) signifies an ordinal/conventional DNA sequence.

Although equivalent to 'CTGATAAC' as an actual DNA sequence expressions, related expressions $\{D_j\}$ and $<D_j>$ differ from each other; the former retains all information regarding inserted 'E's and place numbers whereas the latter does not.

In an extension of the notation, a multiple sequence of deletions of 'E's (say t-times) can be written as a t-tuple of '$<$ $>$'s denoted '$<<<<D_j>>>>$ (t-tuple) $= <D_j>_t$'. The final expression is without explicit 'E's other than those trailing at the end, and thus formulates a genuine DNA sequence after the appearance of indels. (Short for insertion/deletion markers, the idels are strings of mutated base pairs.) Similarly for the operation $\{ \}$, we have '$\{\{\{D_j\}\}\}$ (t-tuple) $= \{D_j\}_t$'. The operations '$\{ \}$' and '$< >$' can be performed freely when necessary; if further indels occur at say 'G_3' and 'A_7' in

$<D_j> = [C_1|T_2|G_3|A_4|T_5|A_6|A_7|C_8|E_9|E_{10}|E_{11}|E_{12}|E_{13}|...]$, then $< D_j >$ changes into $<D_{j1}> = [C_1|T_2(E_3)A_4|T_5|A_6(E_7)C_8|E_9|E_{10}|E_{11}|E_{12}|E_{13}|...]$, and subsequently into $<<D_{j1}>> = [C_1|T_2|A_3|T_4|A_5|C_6|E_7|E_8|E_9|E_{10}|E_{11}|...]$. The sequence $< D_{j1} >$ contains implicit 'E's aside from the trailing 'E's, and can be written as

$\{<D_{j1}>\} = [C_1|T_2|A_4|T_5|A_6|C_8|\mathbf{E_9}|\mathbf{E_{10}}|\mathbf{E_{11}}|\mathbf{E_{12}}|\mathbf{E_{13}}|...]$. Naturally, $\{<D_{j1}>\}$ and $< D_{j1} >$ are equivalent, but $< <D_{j1}> >$ and $< D_{j1} >$ differ. Moreover, as long as place numbers are recognized/traced precisely, combinations of manipulations '$\{ \}$' and '$< >$' are allowed; e.g., $\{<\{\{<D_{j1}>\}\}>\}$. Hence, with appropriate use, we could treat (read, interpret, describe, record) conventional sequences of DNA via '$\{D_j\}$' or '$<D_j>$'. However, below we shall focus on simple sequences 'D_j'.

Looking at the beginning of a base sequence as in the following:

$$D_j = [C_1|G_2|A_3|C_4|...|T_i|...|A_{N-1}|T_N|\mathbf{E_{N+1}}|\mathbf{E_{N+2}}|\mathbf{E_{N+3}}|...],$$

(i: i-th component of D_j, N: the number bases D_j)
a directionality for any D_j can be imposed;

$$D_j(5\rightarrow3) = [C_1|G_2|A_3|C_4|...|T_i|...|A_{N-1}|T_N|\mathbf{E_{N+1}}|\mathbf{E_{N+2}}|\mathbf{E_{N+3}}|...],$$

and

$$D_j(3\rightarrow5) = [T_1|A_2|...|T_{N+1-i}|...|C_{N-3}|A_{N-2}|G_{N-1}|C_N|\mathbf{E_{N+1}}|\mathbf{E_{N+2}}|\mathbf{E_{N+3}}|...].$$

The notation, '$(5 \to 3)$' and '$(3 \to 5)$', is simply an additional label representing the two possible types of endings of single-stranded DNA. Nonetheless, when the number of bases is finite, two sequences can be equivalent, as for example

$$D_j(5 \to 3) = [C_1|G_2|A_3|C_4|T_5|A_6|T_7]$$

and

$$D_j(3 \to 5) = [T_1|A_2|T_3|C_4|A_5|G_6|C_7],$$

unless the prime endings $<5'$(five prime) $\to 3'$(three prime)$>$ or $<3' \to 5'>$ accompanies the sequence designation.

In accordance with these postulates, we now can define the set $D = \{D_j \ (j = 1,2,3,...)|$ $D_j \in C_5 \times C_5 \times C_5 \times ...$ (N times, $N \le \infty$)$\}$ as the set of all possible sequences of recognized N-tuple single-stranded DNA bases. We can regard N to be a positive integer or infinity.

An analogous definition is clearly possible for the set R of RNA sequences; with 'T' substituted by 'U', operations of group Z_5, are similarly definable because all results obtained for DNA pertain to RNA under the base substitution. Thus, set $R = \{R_j \ (j = 1,2,3,...)| \ R_j \in C_5 \times C_5 \times C_5 \times ...$(N times, $N \le \infty$)$\}$ is the set of all possible sequences of recognized N-tuple single-stranded RNA bases with $C_5 = \{C, A, U, G, E\}$.

§2 Group composition that yields changes in DNA bases via a Cartesian vector

Next, we can consider $B = \{B_m \ (m = 1,2,3,...) \ | \ B_m \in Z_5 \times Z_5 \times Z_5 \times ...$(an N-fold product, $N = \infty$)$\} = \{Z_5^{\times N}, \bullet\}$, where elements of B act on any D_j. This means that D_j covers all possible sequences of the DNA bases, and this situation is the same for R_j of sequences of RNA bases.

Because B is a Cartesian product of the same Abelian group, it is also Abelian, where composition of any two elements of B is denoted by '$\bullet$' [4]. Details are shown in Appendix B and Figure 3. Accordingly, its formulation as a group $B = \{Z_5^{\times N}, \bullet\}$ is confirmed.

In a more general context, a Cartesian vector that is composed of the respective operators '$b_{(j \to k)}$' that effects the change D_j into D_k is definable in the following way:

$$B_{(j \to k)} = \left[b_{(j \to k)1}|b_{(j \to k)2}|b_{(j \to k)3}|...|b_{(j \to k)i}|...|b_{(j \to k)(N-1)}|b_{(j \to k)N}|\mathbf{n_{N+1}}|\mathbf{n_{N+2}}|\mathbf{n_{N+3}}|...),$$
$$(N : \text{the number of components}).$$

Hence,

$$\text{'}D_j \circ B_{(j \to k)} = D_k\text{'}. \tag{4}$$

Clearly, for arbitrary 'j' and 'k', there exists a unique 'm' such that '$B_{(j \to k)} = B_m \ (m = 1,2,3,...)$'; despite the difference in notation, the two are identical in practice.

Here, we present a simple example that consists of a multiple product of '$B_{(j \to k)}$'s. Consider the scenario that a certain sequence of a single strand (or one side of a double-strand) of DNA transitions from D_1 to D_3, in stepwise fashion,

$$D_1 = [A_1|C_2|C_3|G_4|T_5|\mathbf{E_6}|\mathbf{E_7}|...] = [A_1|C_2|C_3(\ )G_4|T_5|\mathbf{E_6}|\mathbf{E_7}|...],$$
$$D_2 = [A_1|C_2|C_3(\mathbf{E_4}|\mathbf{E_5})G_6|T_7|\mathbf{E_8}|\mathbf{E_9}|...],$$
$$D_3 = [A_1|C_2|C_3(\mathbf{A_4}|\mathbf{T_5})G_6|T_7|\mathbf{E_8}|\mathbf{E_9}|...],$$
$$D_4 = [C_1|T_2|G_3(T_4|C_5)G_6|A_7|\mathbf{E_8}|\mathbf{E_9}|...].$$

We next consider the change '$D_1 \rightarrow D_2$'. There exists an operator '$B_{(1 \rightarrow 2)} = [n_1|n_2|n_3$ $(\mathbf{r_4}|\mathbf{u_5})l_6|d_7|\mathbf{n_8}|\mathbf{n_9}|...]$' that is able to produce this change, specifically, the insertion of two 'E's between 'C_3' and 'G_4' yields the change '$D_1 \rightarrow D_2$'. However, this sort of manipulation can be troublesome. Hence, in our model, insertion/deletion of 'E's are instead ascribed to the way the vector D_j is interpreted. This is preferable as this avoids easier manipulations. Next, we construct the operator '$B_{(2 \rightarrow 3)}$' that maps '$D_2 \rightarrow D_3$' (the details are shown in Appendix C). With reference to Figures 1, 2 and 3, we find

$$B_{(2 \rightarrow 3)} = [n_1|n_2|n_3(\mathbf{u_4}|\mathbf{d_5})n_6|n_7|\mathbf{n_8}|\mathbf{n_9}|...].$$

In a similar manner,

$$B_{(3 \rightarrow 4)} = [l_1|u_2|d_3(\mathbf{r_4}|\mathbf{d_5})n_6|l_7|\mathbf{n_8}|\mathbf{n_9}|...].$$

Naturally, the final D_4 is obtained from D_1 recursively,

$$\text{'}D_1 \rightarrow D_2\text{'}, \tag{5}$$

$$\text{'}D_2 \circ B_{(2 \rightarrow 3)} \bullet B_{(3 \rightarrow 4)} = D_4\text{'}. \tag{6}$$

From the decomposition

$$\text{'}B_{(j \rightarrow k)} = B_{(j \rightarrow 0)} \bullet B_{(0 \rightarrow k)} = B_j^{-1} \bullet B_k\text{'}, \text{ we obtain } \text{'}D_j \circ B_{(j \rightarrow k)} = D_j \circ B_{(j \rightarrow 0)} \bullet B_{(0 \rightarrow k)}$$
$$= D_0 \bullet B_{(0 \rightarrow k)} = D_k\text{'}, \tag{7}$$

where

$$D_0 = [E_1|E_2|E_3|...|E_i|...|E_{N-1}|E_N|\mathbf{E_{N+1}}|\mathbf{E_{N+2}}|\mathbf{E_{N+3}}|...] \tag{8}$$

denotes the identity element of D.

Note that the group operations can act on D_j irrespective of whether the 'E's are explicit or implicit as defined in §1. Moreover, any sequence 'D_j' can be presented as a polygonal line; as an example, the evolution of changes '$D_1 \rightarrow D_3$' is displayed in Figure 4.
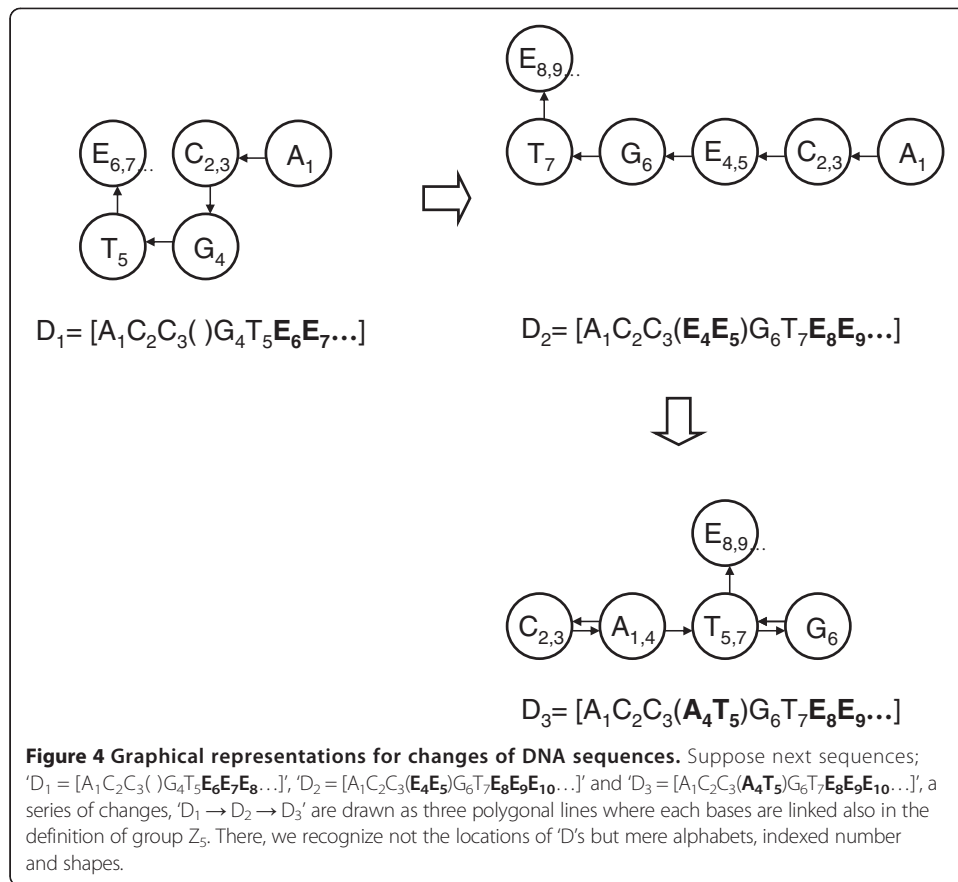
§3 Integration of a linear group and a rotational group as a wallpaper group

Looking at the definitions of groups Z_5, D, and B, another approach is possible. The five bases can be represented by five equispaced phasors with a '$2\pi/5$' angular phase separation located on the unit circle on the Gaussian plane, as depicted in Figure 5.

Herein, in the Gaussian plane, if 'ω' is defined to be the counterclockwise rotational angle '$\omega = 2\pi/5$ (rad)' and composition of 'ω' is denoted '$\bullet$', then assuming 'ω' obeys the 'right translation rule', we have

$$
\begin{aligned}
&\omega = \omega_1, \\
&\omega \bullet \omega = 2\omega = \omega_2, \\
&\omega \bullet \omega \bullet \omega = 3\omega = \omega_3, \\
&\omega \bullet \omega \bullet \omega \bullet \omega = 4\omega = \omega_4, \\
&\omega \bullet \omega \bullet \omega \bullet \omega \bullet \omega = 5\omega = \omega_5 = \omega_0 = 0 \ (= \text{no rotation}).
\end{aligned}
\tag{9}
$$

The general form of an arbitrary base is expressed as '$X_m \leftrightarrow \text{Exp}(m \cdot \omega \cdot i)$' (here, 'i' is the 'imaginary unit', $\omega = 2\pi/5$ (rad), m = {0, 1, 2, 3, 4, 5}). With {*} meaning one of the

Figure 4 Graphical representations for changes of DNA sequences. Suppose next sequences; '$D_1 = [A_1C_2C_3(\ )G_4T_5\mathbf{E_6E_7E_8}...]$', '$D_2 = [A_1C_2C_3(\mathbf{E_4E_5})G_6T_7\mathbf{E_8E_9E_{10}}...]$' and '$D_3 = [A_1C_2C_3(\mathbf{A_4T_5})G_6T_7\mathbf{E_8E_9E_{10}}...]$', a series of changes, '$D_1 \rightarrow D_2 \rightarrow D_3$' are drawn as three polygonal lines where each bases are linked also in the definition of group Z_5. There, we recognize not the locations of 'D's but mere alphabets, indexed number and shapes.

bases among 'C, A, T, G and E', we construct the following map. Denoting composition by '∘', ω_m acts on the identity trivially and hence yields the correspondences

$$Exp(0 \cdot i) \leftrightarrow \{Exp(0 \cdot \omega \cdot i)\} = \{1\} = \{1\} \circ \omega_0 = E = X_0,$$

$$Exp(2\pi i/5) \leftrightarrow \{Exp(1 \cdot \omega \cdot i)\} = \{1\} \circ \omega_1 = C = X_1,$$

$$Exp(4\pi i/5) \leftrightarrow \{Exp(2 \cdot \omega \cdot i)\} = \{1\} \circ \omega_2 = A = X_2,$$

$$Exp(6\pi i/5) \leftrightarrow \{Exp(3 \cdot \omega \cdot i)\} = \{1\} \circ \omega_3 = T = X_3,$$

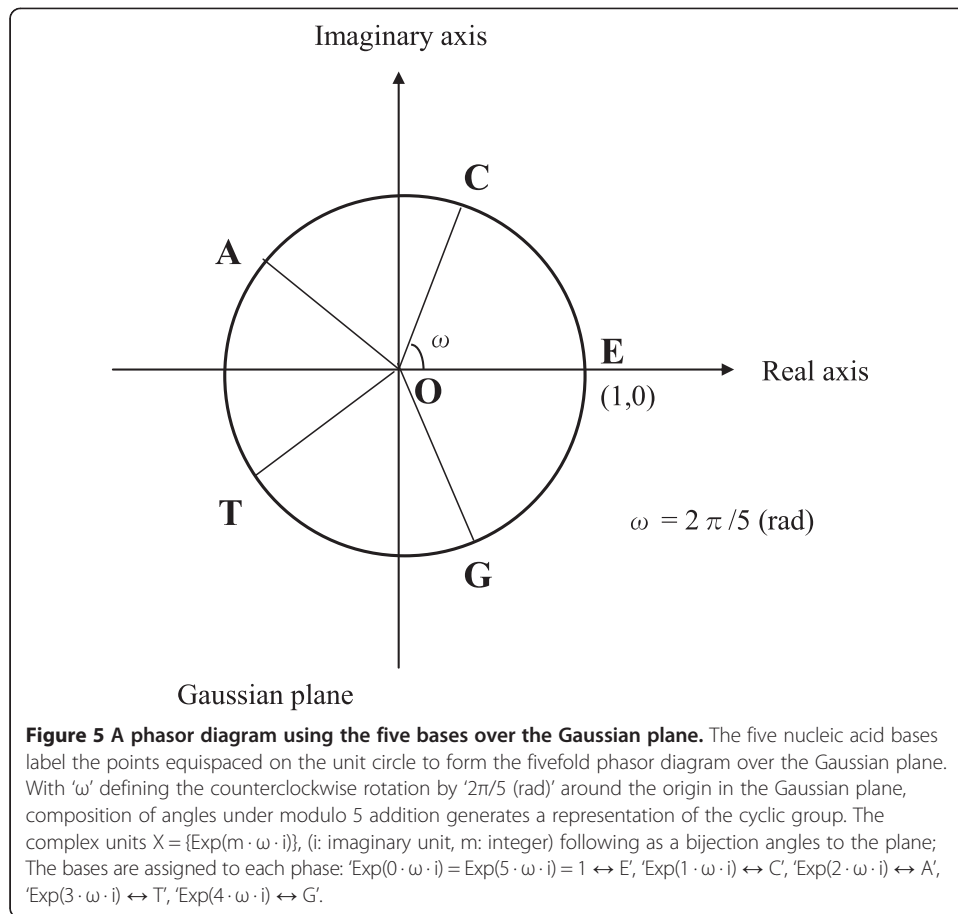$$Exp(8\pi i/5) \leftrightarrow \{Exp(4 \cdot \omega \cdot i)\} = \{1\} \circ \omega_4 = G = X_4,$$

$$X_5 = X_0 = E.$$

$$(10)$$

Expanding the operations for '$\omega_1, \omega_2, \omega_3, ...$' on bases 'C, A, T, G and E', we establish for instance:

$$E \circ \omega_1 = C, C \circ \omega_1 = A, A \circ \omega_2 = G, T \circ \omega_3 = C, G \circ \omega_1 = E.$$

Figure 5 A phasor diagram using the five bases over the Gaussian plane. The five nucleic acid bases label the points equispaced on the unit circle to form the fivefold phasor diagram over the Gaussian plane. With 'ω' defining the counterclockwise rotation by '2π/5 (rad)' around the origin in the Gaussian plane, composition of angles under modulo 5 addition generates a representation of the cyclic group. The complex units X = {Exp(m · ω · i)}, (i: imaginary unit, m: integer) following as a bijection angles to the plane; The bases are assigned to each phase: 'Exp(0 · ω · i) = Exp(5 · ω · i) = 1 ↔ E', 'Exp(1 · ω · i) ↔ C', 'Exp(2 · ω · i) ↔ A', 'Exp(3 · ω · i) ↔ T', 'Exp(4 · ω · i) ↔ G'.

In continuance, the set $P_\omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_0 (= \omega_5)\}$ is readily confirmed to form group $\{P_\omega, \bullet\}$ where the identity element is 'ω_0' and the inverse of 'ω_m' is 'ω_m^{-1}':
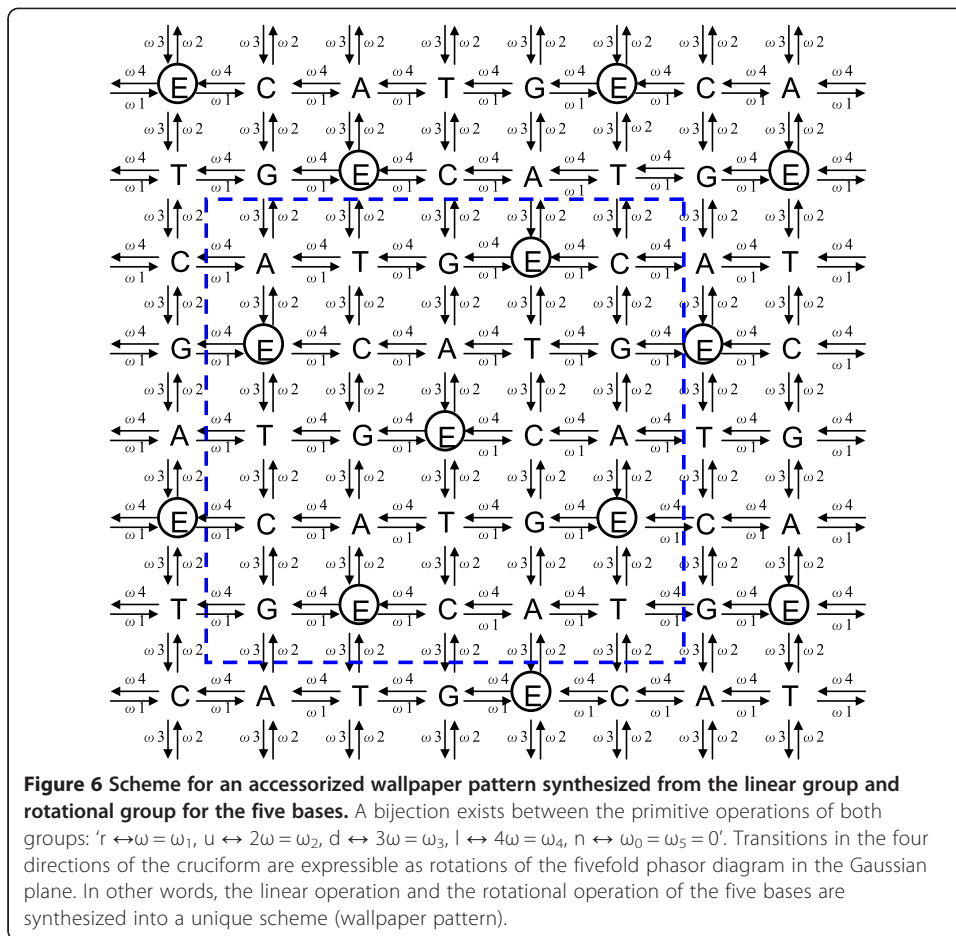
$$
\begin{aligned}
\omega_0{}^{-1} &= \omega_0, \\
\omega_1{}^{-1} &= \omega_4, \\
\omega_2{}^{-1} &= \omega_3, \\
\omega_3{}^{-1} &= \omega_2, \\
\omega_4{}^{-1} &= \omega_1, \\
\omega_5{}^{-1} &= \omega_0 = 0.
\end{aligned}
\tag{11}
$$

Closure and associativity follow from (9) and (10).

Here, if we turn our attention to the wallpaper pattern, a further bijection obeying the postulates of the wallpaper group can be confirmed. Corresponding to Figures 3 and 6 a bijection between the Cayley Tables for translational and rotational operations can be established:

$$
\begin{aligned}
r &\leftrightarrow \omega = \omega_1, \\
u &\leftrightarrow 2\omega = \omega_2, \\
d &\leftrightarrow 3\omega = \omega_3, \\
l &\leftrightarrow 4\omega = \omega_4, \\
n &\leftrightarrow 5\omega = \omega_5 = \omega_0 = 0.
\end{aligned}
\tag{12}
$$

Naturally, inverses (e.g., $\omega_2^{-1} = \omega_3$) are preserved in accordance with the inverses for 'r, u, d, l, and n'. Any right translation of the horizontal line in Figures 1 and 6

Figure 6 Scheme for an accessorized wallpaper pattern synthesized from the linear group and rotational group for the five bases. A bijection exists between the primitive operations of both groups: 'r ↔ ω = ω₁, u ↔ 2ω = ω₂, d ↔ 3ω = ω₃, l ↔ 4ω = ω₄, n ↔ ω₀ = ω₅ = 0'. Transitions in the four directions of the cruciform are expressible as rotations of the fivefold phasor diagram in the Gaussian plane. In other words, the linear operation and the rotational operation of the five bases are synthesized into a unique scheme (wallpaper pattern).

(translational group) is also expressible as a rotation over the fivefold phasor diagram in Figure 5 (rotational group). Thus, these are able to be regarded as a synthesized form of the wallpaper style (wallpaper group) from which expressions such as 'A∘r = A∘ω₁ = T = E∘d' and 'A∘l = A∘ω₄ = C = E∘r' can be confirmed. All possible one-step changes between 'A, C, T, G and E' and 'ω₁, ω₂, ω₃, ω₄ and ω₀,' are shown in Figure 2.

Therefore, this rule for 'E' does not break the postulates for set D, group Z_5, and group B.

§4 Methods to obtain complementary sequences from primary DNA

Suppose, from among 'C, A, T, G and E', a base 'X_m' is given; its complementary base '$X_m^\dagger$' to 'X_m' is defined as follows; for '$X_m = \{Exp(m \cdot \omega \cdot i)\}$, m = \{0, 1, 2, 3, 4, 5\}', then '$X_m^\dagger$' is obtained by '$X_m^\dagger = \{Exp((5 - m) \cdot \omega \cdot i)\}$', where '$\{Exp(5 \cdot \omega \cdot i)\} = \{1\} = E$'. In this regard,

$$\text{'}X_5^\dagger = X_0 = X_0^\dagger = X_5\text{'} \tag{13}$$

The procedure yields specifically '$A^\dagger = T$' and '$C^\dagger = G$'.
Clearly, the complement of 'E' is 'E' itself; '$E^\dagger = E$'.

Another notation for the 'X_m' expressed as a base can be given. We introduce the one-value function '$^\omega X(m)$' that provides the same results,

$$'X_m = {}^\omega X(m) = E \cdot m\omega = E \circ (\omega \bullet \omega \bullet \ldots \bullet \omega \ (m \ times))'(m = 0, 1, 2, \ldots, 5). \tag{14}$$

As for 'm' in (14), both positive and negative integers are permissible. Thus, '$X_m^\dagger$' is expressible as

$$X_m^\dagger = {}^\omega X(5-m) = E \cdot (5-m)\omega = E \circ (\omega \bullet \omega \bullet \ldots \bullet \omega \ ('5-m' \ times))'(m = 0, 1, 2, 3, 4, 5). \tag{15}$$

A simple example is illustrated below.
Suppose 'D_j' = $[A_1|T_2|C_3|E_4|G_5|T_6|\ldots] = [{}^\omega X(2)|{}^\omega X(3)|{}^\omega X(1)|{}^\omega X(0)|{}^\omega X(4)|{}^\omega X(3)|\ldots]$, then,

$$
\begin{aligned}
'D_j^\dagger' &= [{}^\omega X(5-2)|{}^\omega X(5-3)|{}^\omega X(5-1)|{}^\omega X(5-0)|{}^\omega X(5-4)|{}^\omega X(5-3)|\ldots], \\
&= [{}^\omega X(3)|{}^\omega X(2)|{}^\omega X(4)|{}^\omega X(0)|{}^\omega X(1)|{}^\omega X(2)|\ldots], \\
&= [T_1|A_2|G_3|E_4|C_5|A_6|\ldots].
\end{aligned}
\tag{16}
$$

In accordance with the wallpaper group in Figure 1, the translations in one direction (e.g., right) over a horizontal line form a cyclic group P_r that contains only $\{r, r^2, r^3, r^4, r_e \ (= r^0 = r^5 = n)\}$. This group is isomorphic with group $P_\omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_0 \ (= \omega_5)\}$, as is the group similarly generated over a vertical line.

Similar to '$^\omega X(m)$', 'X_m' can be expressed using another one-value function $^r X(m) = E \circ r^m$:

$$X_m = {}^r X(m) = E \cdot r^m = E \circ (r \bullet r \bullet \ldots \bullet r \ (m \ times)) \ (m = 0, 1, 2, 3, 4, 5). \tag{17}$$

Hence, '$X_m^\dagger$' (the complementary base of 'X_m') is written as

$$X_m^\dagger = {}^r X(5-m) = E \cdot r^{5-m} = E \circ (r \bullet r \bullet \ldots \bullet r \ ('5-m' \ times)). \tag{18}$$

Extension to vertical translations is straightforward;

$$'X_m = {}^u X(5) = E \cdot u^5 = E \circ (u \bullet u \bullet \ldots \bullet u \ (m \ times)), \tag{19}$$

and its complementary base '$X_m^\dagger$' can be identified similarly although the order of letters are somewhat different.

Consider the following simple example in identifying '$D_j^\dagger$' using '$^r X(m)$'s;

$$\text{for } 'D_j' = [A_1|T_2|C_3|E_4|G_5|T_6|\ldots] = [{}^r X(2)|{}^r X(3)|{}^r X(1)|{}^r X(0)|{}^r X(4)|{}^r X(3)|\ldots],$$

by replacing '$^\omega X(m)$' by '$^r X(m)$' in formula (15), the same result is obtained.

According to these rules, '$X_1 \circ b_{[X1 \to X4]} = X_1 \circ r^3 = E \circ r^4 = G$'. In general, when the i-th component '$b_{[Xm1 \to Xm2]i}$' of '$B_{(m1 \to m2)}$' changes X_{m1} ($= E \circ r^{m1} = E \circ (m_1\omega)$) to X_{m2} ($= E \circ r^{m2} = E \circ (m_2\omega)$).

Hence, the highlighted form of the operator vector is expressed as

$$B_{(m1 \to m2)} = [...|b_{[Xm1 \to Xm2]\,i}|...] = [...|r^{m2-m1}{}_i|...] = [...|(m_2-m_1)\omega_i|...]. \tag{20}$$

For a further example, given the operator '$B_{(j \to k)}$' that changes D_j to D_k,

$$D_j = [C_1|A_2|E_3|...|C_i|...|T_{N-1}|A_N|\mathbf{E_{N+1}}|\mathbf{E_{N+2}}|\mathbf{E_{N+3}}|...],$$
$$= [E{\cdot}r^1{}_1|E{\cdot}r^2{}_2|E{\cdot}r^0{}_3|...|E{\cdot}r^1{}_i|...|E{\cdot}r^3{}_{N-1}|E{\cdot}r^2{}_N|\mathbf{E{\cdot}r^0{}_{N+1}}|\mathbf{E{\cdot}r^0{}_{N+2}}|\mathbf{E{\cdot}r^0{}_{N+3}}|...],$$
$$D_k = [G_1|T_2|C_3|...|G_i|...|C_{N-1}|T_N|\mathbf{E_{N+1}}|\mathbf{E_{N+2}}|\mathbf{E_{N+3}}|...],$$
$$= [E{\cdot}r^4{}_1|E{\cdot}r^3{}_2|E{\cdot}r^1{}_3|...|E{\cdot}r^4{}_i|...|E{\cdot}r^1{}_{N-1}|E{\cdot}r^3{}_N|\mathbf{E{\cdot}r^0{}_{N+1}}|\mathbf{E{\cdot}r^0{}_{N+2}}|\mathbf{E{\cdot}r^0{}_{N+3}}|...].$$

With details shown in Appendix D, '$B_{(j \to k)}$' takes the form

$$B_{(j \to k)} = [r^3{}_1|r^1{}_2|r^1{}_3|...|r^3{}_i|...|r^3{}_{N-1}|r^1{}_N|\mathbf{r^0{}_{N+1}}|\mathbf{r^0{}_{N+2}}|\mathbf{r^0{}_{N+3}}|...].$$

Naturally, the state D_k is obtained through recursively applying the operations, $D_j \circ B_{(j \to k)} = D_k$. (Details are presented in Appendix D).

Whereas '$D_j{}^\dagger$'s might have components in reverse order in terms of sense (5' or 3'), there exists however certain 'D_k' such that '$D_k = D_j{}^\dagger$, (j, k = 0, 1, 2, 3, 4,...)'. With this, '$D_j{}^\dagger$' is one of the ordinal elements belonging to the same set D. Thus, the symbol '†' need only be present when elements are distinct.

§5 Further unifying notation to describe the wallpaper group operation

Consider Figure 6; we assume that the number of right translations 'r' $\in$ group P_r, (or ω $\in$ group P_r) is 'a' and the number of up translations 'u' $\in$ group P (or 2ω (= ω_2) $\in$ group P_r) is 'b' with a, b = ...,-2, –1, 0, 1, 2,.... Similarly, with 'd $\leftrightarrow$ 3ω' 'l $\leftrightarrow$ 4ω', the total change can be summarized as 'x[a, b]'. We can confirm that there exists at least a pair of 'a, b' that satisfies

$$\text{'X} = E{\cdot}x[a, b]\text{'}, \tag{21}$$

because any base in Figure 6 can be obtained by a finite number of transitions from 'E'. For instance, A can be expressed as; 'A = E∘x[0, 1] = E∘x[2, 0] = E∘x[1,3] = E∘x[2,3] = E∘x [4,4]'. However, we remark that 'x[a, b]' means changes of bases from one to another prescribed by the wallpaper pattern. In practice, 'x[a, b] = $r^a \bullet u^b$' constitutes a multiple composition of elements of group Z_5. In addition, 'x[–a, –b] = $r^{-a} \bullet u^{-b}$ = $l^a \bullet d^b$' or $\leftrightarrow$ '(–a)$\omega \bullet$(–b)(2ω) = (–a–2b)ω'. E.g.,'x[–3, –2] means '$r^{-3} \bullet u^{-2}$ = $l^3 \bullet d^2$' or $\leftrightarrow$ '(–3)$\omega \bullet$(–2·2) (ω) = (–3 -4)ω = (–7)ω = (–2)ω = 3ω = ω_3'.

For the wallpaper group, the 'a' and 'b' should be interpreted in modulo 5 addition. The Cayley table for the wallpaper group are presented in Appendix A.

Within 'the square unit cell' in Figure 1 or 6, there are five pairs of 'a, b' for each base, as for 'A'. Under modulo 5 addition, 'x[a + 5, b + 5] = x[a, b]' holds. Moreover, if '$X^\dagger$' is obtained from 'X' using 'X = E∘x[a, b]', '$X^\dagger$' can be determined as

$$\text{'}X^\dagger = E{\cdot}x^{-1}[a, b] = E{\cdot}x[-a, -b]\text{'}, \tag{22}$$

$$\text{or '}X^\dagger = E{\cdot}x[5-a, 5-b]\text{'}, \tag{23}$$

because 'X' and '$X^\dagger$' are symmetrically disposed with respect to 'E' over the wallpaper pattern that would be selected as a standard for the definition of 'a, b'. In practice, for

an arbitrary 'X', 'X†' can be obtained via (22) or (23) by making use of an arbitrary 'E' as the reference point for the symmetry.

For example, if 'G = E∘x[1,2]', then according to (22) 'G† = E∘x^{-1}[1,2] = E∘x[−2, −1] = C' or according to (23) G† = E∘x[1,2] = E∘x[3,4] = C'. There are an infinite number of identifications for the complementary base for an arbitrary base 'X'.

Moreover, if we define the composition for the 'x[a, b]'s as

$$\text{'}x[a_1, b_1] \bullet x[a_2, b_2] = x[a_1, b_1] + x[a_1, a_2] = x[a_1 + a_2, b_1 + b_2]\text{'}, \tag{24}$$

we can confirm descriptions (20), (22) and (23). As for the operators 'B_m', with 'D_j' expressed as '$D_j = [...|E\circ x[a_i \ , b_i] \ _i|...]$', one of the candidates of the appropriate '$B_{(j\to j\dagger)}$'s that produces '$D_j \circ B_{(j\to j\dagger)} = D_j^{\dagger}$' is identified:

$$B_{(j\to j\dagger)} = \left[...\left|x\left[-2a_i, -2b_{ii}\right|...\right] = \left[...\left|r^{-2ai}\bullet u^{-2bi}\right._i|...\right] = \left[...\left|l^{2ai}\bullet d^{2bi}\right._i|...\right], \tag{25a}$$

$$\text{or} = \left[...\left|r^{-2ai}\bullet\left(r^2\right)^{-2bi}\right._i|...\right] = \left[...\left|r^{-2ai-4bi}\right._i|...\right] = \left[...\left|l^{2ai+4bi}\right._i|...\right] \ \left(\text{using '}u = r^2\text{', 'd = }l^2\text{'}\right), \tag{25b}$$

$$\text{or}\leftrightarrow\left[...\left|(-2a_i)\omega\bullet(-2b_i)\cdot(2\omega)\right._i|...\right] = \left[...\left|(-2a_i-4b_i)\omega_i\right|...\right] \ \left(\text{using '}r^m\leftrightarrow m\omega\text{', 'u}^m\leftrightarrow 2m\omega\text{'}\right). \tag{25c}$$

The exponents '-2a_i -4b_i' in (25b) are permitted to take positive or negative integer values.

In these expressions, the rules for the wallpaper group (25a) can also be expressed as either for the linear group or for the rotational group (25b or 25c).

More generally,

'$D_j = [...|E\circ x[a_{(j)i} \ , b_{(j)i}] \ _i|...]$' is changed into '$D_k = [...|E\circ x[a_{(k)i} \ , b_{(k)i}] \ _i|...]$', and '$B_{(j\to k)}$'s that provides '$D_j \circ B_{(j\to k)} = D_j^{\dagger}$' is identified as

$$B_{(j\to k)} = \left[...\left|xa_{(k)i}-a_{(j)i}, b_{(k)i}-b_{(j)ii}\right|...\right] = \left[...\left|r^{a(k)i-a(j)i}\bullet u^{b(k)i-b(j)i}\right._i\right)...\right)$$

$$\left(= \left[...\left|l^{-a(k)i+a(j)i}\bullet d^{-b(k)i+b(j)i}\right._i|...\right]\right). \tag{26a}$$

$$\text{Also,} = \left[...\left|r^{a(k)i-a(j)i}\bullet\left(r^2\right)^{b(k)i-b(j)i}\right._i|...\right] = \left[...\left|r^{a(k)i-a(j)i+2b(k)i-2b(j)i}\right._i|...\right], \tag{26b}$$

$$\text{or else}\leftrightarrow\left[...\left|\left(a_{(k)i}-a_{(j)i}\right)\omega\bullet 2\left(b_{(k)i}-b_{(j)i}\right)\omega_i\right|...\right] = \left[...\left|\left(a_{(k)i}-a_{(j)i} + 2b_{(k)i}-2b_{(j)i}\right)\omega_i\right|...\right]. \tag{26c}$$

As mentioned in §1, if a certain sequence 'X' has sense <5' → 3'>, the complementary sequence 'X†' of a certain sequence 'X' is reversed to <3' → 5' > .

To aid understanding, we present the following examples: Given

$$D_j = [A_1|T_2|E_3|...|C_i|...|G_{N-1}|A_N|\mathbf{E_{N+1}}|\mathbf{E_{N+2}}||\mathbf{E_{N+3}}|...]$$
$$= [E \circ x[0, \ 1]_1|E \circ x[0, \ -1]_2||E \circ x[0, \ 0]_3|...|E \circ x[1, \ 0]_i|...$$
$$...|E \circ x[-1, \ 0]_{N-1}|E \circ x[0, \ 1]_N|\mathbf{E \circ x[0, \ 0]_{N+1}}|\mathbf{E \circ x[0, \ 0]_{N+2}}|\mathbf{E \circ x[0, \ 0]_{N+3}}|...].$$

then, according to (22), '$D_j^{\dagger}$' is simply

$$D_j{}^{\dagger} = \left[E \circ x^{-1}[0,1]_1 \middle| E \circ x^{-1}[0,-1]_2 \middle| E \circ x^{-1}[0,0]_3 \middle| ...\right.$$
$$...|E \circ x^{-1}[1,0]_i|...|E \circ x^{-1}[-1,0]_{N-1}|E \circ x^{-1}[0,1]_N|\mathbf{E \circ x^{-1}[0,0]_{N+1}}|\mathbf{E \circ x^{-1}[0,0]_{N+2}}|\mathbf{E \circ x^{-1}[0,0]_{N+3}}|...],$$
$$= [E \circ x0,-1]_1|E \circ x[0,1]_2|E \circ x[0,0]_3|...$$
$$...|E \circ x[-1,0]_i|...|E \circ x[1,0]_{N-1}|E \circ x[0,-1]_N|\mathbf{E \circ x[0,0]_{N+1}}|\mathbf{E \circ x[0,0]_{N+2}}|\mathbf{E \circ x[0,0]_{N+3}}|...],$$
$$= [T_1|A_2|E_3|...|G_i|...|C_{N-1}|T_N|\mathbf{E_{N+1}}|\mathbf{E_{N+2}}|\mathbf{E_{N+3}}|...].$$

If we use the optional formula (25a – c), the relation '$D_j \circ B_{(j \to jt)} = D_j{}^{\dagger}$' is derived. Details are given in Appendix E.

Apart from these examples, additional identities for the wallpaper group can be verified using Figure 1 or 6; e.g.,

'$x[1, 0] \bullet x[0, 1] = x[1, 0] + x[0, 1] \ (= r \bullet u) = x[1 + 0, 0 + 1] = x[1, 1] = x[0, -1] (= d)$',
'$x[2, 0] (= r \bullet r) = x[0, 1] = u$'.'$x[3, 1] \bullet x[-1, 10] = x[3-1, 1 + 10] = x[2, 11] = x[2, 1]$
$= r^2 \bullet u^1 = (r \bullet r) \bullet u = u \bullet u = l$'.

We develop various general formulas:

'$x[a + 5, \ b] = x[a, \ b]$', '$x[a, \ b + 5] = x[a, \ b]$',
'$x[2a, \ -a] = x[0, \ 0]$', '$x[a, \ 2a] = x[0,0]$', $\qquad\qquad$ (27)
'$x[-2a, \ a] = x[0, \ 0]$', '$x[-a, \ -2a] = x[0, \ 0]$'.

Other unknown rules might underlie the wallpaper pattern.

Concerning style in treating the wallpaper group, examples '$X_m = {}^r X(m)$ in (16, 17, 19), and $^{\omega}X(m)$ in (14, 15, 20) could be regarded as a specific combination that are displayed as

$$'X_m = {}^r X(m) = {}^{\omega} X(m) = E \circ x[a, 0] \ (a = ..., -2, -1, 0, 1, 2, ..., 5, 6, ...; \text{integer})'. \qquad (28)$$

§6 Treatment of changes of sequences and the insertion/deletion of DNA bases via an optionally generalized operation

Below, we demonstrate, using several examples containing 'E's, changes and inclusion/exclusion of DNA bases using a more generalized scheme.

For definiteness, let 'D_j' be the sequence '**CGT**AT...C...TA', we consider the change of its '1–3' components '$\mathbf{C_1G_2T_3}$' into '$\mathbf{G_1T_2A_3}$', and moreover the insertion of two bases 'GC' between 'T_3' and 'A_4' denoted '( )':

$$D_j = \left[\mathbf{C_{(j)1}} \middle| \mathbf{G_{(j)2}} \middle| \mathbf{T_{(j)3}}(\ )A_{(j)4} \middle| ... \middle| C_{(j)i} \middle| ... \middle| T_{(j)N-1} \middle| A_{(j)N} \middle| \mathbf{E_{(j)N+1}} \middle| \mathbf{E_{(j)N+2}} \middle| \mathbf{E_{(j)N+3}} \middle| ...\right].$$

We denote the result of this transformation as D_k,

$$D_k = [\mathbf{G_{(k)1}} \middle| \mathbf{T_{(k)2}} \middle| \mathbf{A_{(k)3}}(\mathbf{G_{(k)4}} \middle| \mathbf{C_{(k)5}})A_{(k)6} \middle| ... \middle| C_{(k)i+2} \middle| ... \middle| T_{(k)N+1} \middle| A_{(k)N+2} \middle| \mathbf{E_{(k)N+3}} \middle| \mathbf{E_{(k)N+4}} \middle|$$
$$\mathbf{E_{(k)N+5}}... \middle| ...\right].$$
$$\qquad\qquad (29)$$

The procedure from D_j to D_k is described recursively to find operator '$B_{(j \to h)}$'.

First, two 'E's are inserted after the 3rd component (this change is denoted '$D_j \to D_h$') in preparation for insertion of 'GC';

$$D_j \to D_h = [\mathbf{C_{(h)1}}|\mathbf{G_{(h)2}}|\mathbf{T_{(h)3}}(\mathbf{E_{(h)4}}|\mathbf{E_{(h)5}})A_{(h)4+2}|...|C_{(h)i+2}|...$$
$$...|T_{(h)N-1+2}|A_{(h)N+2}|\mathbf{E_{(h)N+1+2}}|\mathbf{E_{(h)N+2+2}}|\mathbf{E_{(h)N+3+2}}|...]. \tag{30}$$

Thus, '$B_{(j \to h)}$

$$= [b_{[C \to C]1}|b_{[G \to G]2}|b_{[T \to T]3}(b_{[G \to E]4}|b_{[C \to E]5})b_{[A \to A]6}|...|b_{[C \to C]i}|...|b_{[T \to T]N-1}|b_{[A \to A]N}]'.$$

This change is in accordance with those rules for vector-like 'D_j's dependent upon 'E's.

Hence, the operator $B_{(h \to k)}$ that produces the change from D_h to D_k is described as:

$$B_{(h \to k)} = [b_{[C \to G]1}|b_{[G \to T]2}|b_{[T \to A]3}(\mathbf{b_{[E \to G]4}}|\mathbf{b_{[E \to C]5}})b_{[A \to A]6}|...$$
$$...|b_{[C \to C]i+2}|...|b_{[T \to T]N+1}|b_{[A \to A]N+2}|\mathbf{b_{[E \to E]N+3}}|\mathbf{b_{[E \to E]N+4}}|\mathbf{b_{[E \to E]N+5}}|...]. \tag{31}$$

Thereby,

$$D_h \circ B_{(h \to k)}$$
$$= [C_1 \circ b_{[C \to G]1}|G_2 \circ b_{[G \to T]2}|T_3 \circ b_{[T \to A]3}(\mathbf{E_4} \circ \mathbf{b_{[E \to G]4}}|\mathbf{E_5} \circ \mathbf{b_{[E \to C]5}})A_6 \circ b_{[A \to A]6}|...|C_i \circ b_{[C \to C]i+2}|...$$
$$|T_{N+1} \circ b_{[T \to T]N+1}|A_{N+2} \circ b_{[A \to A]N+2}|\mathbf{E_{N+3}} \circ \mathbf{b_{[E \to E]N+3}}|\mathbf{E_{N+4}} \circ \mathbf{b_{[E \to E]N+4}}|\mathbf{E_{N+5}} \circ \mathbf{b_{[E \to E]N+5}}|...].$$
$$\tag{32}$$

With reference to Figure 1, 2, 6 or Appendix A,

$$= [C_1 \circ d_1|G_2 \circ l_2|T_3 \circ l_3(\mathbf{E_4} \circ \mathbf{l_4}|\mathbf{E_5} \circ \mathbf{r_5})A_6 \circ n_6|...|C_{i+2} \circ n_{i+2}|...$$
$$...|T_{N+1} \circ n_{N+1}|A_{N+2} \circ n_{N+2}|\mathbf{E_{N+3}} \circ \mathbf{n_{N+3}}|\mathbf{E_{N+4}} \circ \mathbf{n_{N+4}}|\mathbf{E_{N+5}} \circ \mathbf{n_{N+5}}|...], \tag{33}$$

$$= [G_1|T_2|A_3(\mathbf{G_4}|\mathbf{C_5})A_6|...|C_{i+2}|...|T_{N+1}|A_{N+2}|\mathbf{E_{N+3}}|\mathbf{E_{N+4}}|\mathbf{E_{N+5}}|...], \tag{34}$$
$$= D_k.(29).$$

This indicates a code change of the '1–3' components and a 'GC' insertion after the 3rd as described via the two steps: 1) $D_j \to D_h$ (inserting two 'E's after the '3rd' component), and 2) $D_h \circ B_{(h \to k)} = D_k$. Note that the exclusion of the '4–5' components 'GC' from D_k and the transformation of the '1–3' components from 'GTA' to 'CGT' constitute the recursive procedure for the inverse operator

$$'B_{(k \to h)} = B_{(h \to k)}{}^{-1}' \tag{35}$$

Alternatively, '$D_h \to D_j$' is obtained by deleting the two 'E's from the '4–5' components of D_h to yield the initial state 'D_j' in accordance with the characteristics of the vector-like 'D_j's.

In summary, essentially, all transitions (changes and inclusion/exclusion) of a certain sequence within the same single-stranded DNA, whether it has finite or infinite length, can be described in principle within a single operation using only the unique operator $B_{(... \to ...)} \in$ group B.

§7 Synthesis of changes, insertion/deletion, and recombination of DNA bases

As a further development, to demonstrate recombination, take two finite sequences 'GETAGT (= D_{c1})' and 'ATAGCTA (= D_{d1})'. These have vector expressions

$$D_{c1} = [G_1 E_2 \underline{T_3 A_4 G_5 T_6} \,|\mathbf{E_7 E_8 E_9}...], \tag{36}$$

$$= [G_{(c1)1}|E_{(c1)2}|T_{(c1)3}|A_{(c1)4}|G_{(c1)5}|T_{(c1)6}|\mathbf{E_{(c1)7}}|\mathbf{E_{(c1)8}}|\mathbf{E_{(c1)9}}|...],$$

$$D_{d1} = [A_1 T_2 \underline{A_3 G_4 C_5 T_6 A_7} \,|\mathbf{E_8 E_9 E_{10}}...], \tag{37}$$

$$= \left[A_{(d1)1}|T_{(d1)2}|A_{(d1)3}|G_{(d1)4}|C_{(d1)5}|T_{(d1)6}|A_{(d1)7}|\mathbf{E_{(d1)8}}|\mathbf{E_{(d1)9}}|\mathbf{E_{(d1)10}}|...\right].$$

To illustrate for the pair D_{c1} and D_{d1}, we consider recombination to take place between the sequence '$T_3 A_4 G_5 T_6$' of the (3–6)-th component of 'D_{c1}' and the 'AGCTA' of the (3–7)-th component of 'D_{d1}' at the same instant.

First, in the pair of sequences, a series of 'E's of complementary size is inserted in 'D_{c1}' just before the sequence to be converted, and in 'D_{d1}' just after the sequence to be converted. For example, for 'D_{c1}', five 'E's, '**EEEEE**', of size equivalent to that of '$\mathbf{A_3 G_4 C_5 T_6 A_7}$' of '$D_{d1}$', are inserted just before 'T_3' in 'D_{c1}' where '$\mathbf{A_3 G_4 C_5 T_6 A_7}$' is to be located, that is, the interval between 'the 2nd 'E_2' and 3rd 'T_3' within 'D_{c1}'. Under this procedure, D_{c1} changes into D_{c2}:

$$D_{c2} = [G_1 E_2 (\mathbf{EEEEE}) \underline{T_{3+5} A_{4+5} G_{5+5} T_{6+5}} \,|\mathbf{E_{7+5} E_{8+5} E_{9+5}}...], \tag{38}$$

$$= [G_1 E_2 (\mathbf{EEEEE}) \underline{T_8 A_9 G_{10} T_{11}} |\mathbf{E_{12} E_{13} E_{14}}...].$$

Here, we assume that '**EEEEE**' is changed into '$\mathbf{A_3 G_4 C_5 T_6 A_7}$' (originally, the (3–7)-th component of 'D_{d1}'). In addition, '$\underline{T_8 A_9 G_{10} T_{11}}$' is transformed into the same number of 'E's, '$\underline{EEEE}$', at the same time. By this process, 'D_{c2}' changes in 'D_{c3}':

$$D_{c3} = \left[G_1 E_2 (\mathbf{A_3 G_4 C_5 T_6 A_7}) \underline{E_8 E_9 E_{10} E_{11}} |\mathbf{E_{12} E_{13} E_{14}}...\right]. \tag{39}$$

Note that bold type and underline are here merely pedagogical aids to help identify sequence changes. Meanwhile, four 'E's '**EEEE**' equivalent in size to '$\underline{T_3 A_4 G_5 T_6}$' of '$D_{c1}$' would be inserted after 'A_7' of 'D_{d1}' where '$\underline{T_3 A_4 G_5 T_6}$' of '$D_{c1}$' is to be located within 'D_{d1}'. That is, '$\underline{T_3 A_4 G_5 T_6}$' is inserted into the interval between the 7th 'A_7' and 8th 'E_8' within 'D_{d1}'. In this procedure, D_{d1} changes into D_{d2}:

$$D_{d2} = \left[A_1 T_2 \underline{A_3 G_4 C_5 T_6 A_7} (\mathbf{EEEE}) \mathbf{E_{8+4} E_{9+4} E_{10+4}}...\right],$$

$$= \left[A_1 T_2 \underline{A_3 G_4 C_5 T_6 A_7} (\mathbf{EEEE}) \mathbf{E_{12} E_{13} E_{14}}...\right]. \tag{40}$$

Furthermore, we change 'EEEE' into the equivalent-sized '$\mathbf{T_8 A_9 G_{10} T_{11}}$' (originally, the (3-6)-th components of 'D_{c1}') while '$\underline{A_3 G_4 C_5 T_6}\ \underline{A_7}$' is transformed into the equivalent-sized 'EEEEE'. Through this procedure, 'D_{d2}' changes in 'D_{d3}':

$$D_{d3} = \left[\underline{A_1 T_2 E_3 E_4 E_5 E_6 E_7} (\mathbf{T_8 A_9 G_{10} T_{11}}) E_{12} E_{13} E_{14}...\right] \tag{41}$$

As a result, if we omit the infinite series of 'E's from right end, we have the recombination (partial conversion between this pair of sequences from 'D_{c1}, D_{d1}') with 'D_{c1}' = '$G_1 E_2 \underline{T_3 A_4 G_5 T_6}$' being transformed into 'D_{c3}' = '$G_1 E_2 \underline{A_3 G_4 C_5 T_6 A_7}$' and '$D_{d1}$' = '$A_1 T_2 \underline{A_3 G_4 C_5 T_6 A_7}$' being transformed into 'D_{d3}' = '$A_1 T_2 \underline{T_3 A_4 G_5 T_6}$'. We

define the manipulation of the recombination (partial/total conversion) between 'D_{c1}, D_{d1}' in this way.

In the initial stage in the previous illustration, we inserted different sizes of 'E' sequences in each line; however, processes '$D_{c1} \rightarrow D_{c2}$' and '$D_{d1} \rightarrow D_{d2}$' are preferred to be regarded as 'E' insertions/deletions (see comments prior to equation (5)) and this rule depends upon the characteristics of these vectors (e.g., 'D_j's).

As previously explained, the operations can be performed in any of the three equivalent linear group, rotational group, and wallpaper group. Choosing the wallpaper group,

$$
\begin{aligned}
&D_{c2} \circ B_{(c2 \rightarrow c3)} \\
&= [G_1 \circ n_1 | E_2 \circ n_2 (\mathbf{E_3} \circ \mathbf{u_3} | \mathbf{E_4} \circ \mathbf{l_4} | \mathbf{E_5} \circ \mathbf{r_5} | \mathbf{E_6} \circ \mathbf{d_6} | \mathbf{E_7} \circ \mathbf{u_7}) \underline{T_8 \circ u_8} \ | \underline{A_9 \circ d_9} | \underline{G_{10} \circ r_{10}} | \underline{T_{11} \circ u_{11}} | \mathbf{E_{12}} \circ \mathbf{n_{12}} \\
&\quad | \mathbf{E_{13}} \circ \mathbf{n_{13}} | \mathbf{E_{14}} \circ \mathbf{n_{14}} | ...],
\end{aligned}
\tag{42}
$$

$$
= [G_1 | E_2 (\mathbf{A_3} | \mathbf{G_4} | \mathbf{C_5} | \mathbf{T_6} | \mathbf{A_7}) \underline{E_8} | \underline{E_9} | \underline{E_{10}} | \underline{E_{11}} | \mathbf{E_{12}} | \mathbf{E_{13}} | \mathbf{E_{14}} | ...],
$$

$$
= D_{c3},
$$
$$
\text{where } B_{(c2 \rightarrow c3)} = [n_1 | n_2 (\mathbf{u_3} | \mathbf{l_4} | \mathbf{r_5} | \mathbf{d_6} | \mathbf{u_7}) \underline{u_8} | \underline{d_9} | \underline{r_{10}} | \underline{u_{11}} | \mathbf{n_{12}} | \mathbf{n_{13}} | \mathbf{n_{14}} | ...].
\tag{43}
$$

Also,

$$
\begin{aligned}
&D_{d2} \circ B_{(d2 \rightarrow d3)} \\
&= [A_1 \circ n_1 | T_2 \circ n_2 | \underline{A_3 \circ d_3} \ | \underline{G_4 \circ r_4} \ | \underline{C_5 \circ l_5} \ | \underline{T_6 \circ u_6} | A_7 \circ d_7 \\
&\quad (\mathbf{E_8} \circ \mathbf{d_8} | \mathbf{E_9} \circ \mathbf{u_9} | \mathbf{E_{10}} \circ \mathbf{l_{10}} | \mathbf{E_{11}} \circ \mathbf{d_{11}}) E_{12} \circ n_{12} | E_{13} \circ n_{13} | E_{14} \circ n_4 | ...],
\end{aligned}
\tag{44}
$$

$$
= [A_1 | T_2 | \underline{E_3} | \underline{E_4} | \underline{E_5} | \underline{E_6} | \underline{E_7} (\mathbf{T_8} | \mathbf{A_9} | \mathbf{G_{10}} | \mathbf{T_{11}}) E_{12} | E_{13} | E_{14} | ...],
$$

$$
= D_{d3},
$$
$$
\text{where } B_{(d2 \rightarrow d3)} = [n_1 | n_2 | \underline{d_3} | \underline{r_4} | \underline{l_5} | \underline{u_6} | \underline{d_7} (\mathbf{d_8} | \mathbf{u_9} | \mathbf{l_{10}} | \mathbf{d_{11}}) n_{12} | n_{13} | n_{14} | ...].
\tag{45}
$$

With respect to (42) and (44), the inverse identities are confirmed:

$$
B_{(c2 \rightarrow c3)}{}^{-1} = B_{(d2 \rightarrow d3)}.
\tag{46}
$$

Generally, $B_{(\_ \rightarrow \_)}$ giving transition '$D_{c2} \rightarrow D_{c3}$' automatically produces an inverse change for '$D_{d2} \rightarrow D_{d3}$', as stated in (46) and reduces troublesome manipulations, even if only partially.

§8 Further applications of the composition category-like prototypal model using additional ribonucleic acid (RNA)

We next comment on other possible applications of the model. The category theory-like construction for treating DNA transcription to RNA might be conceivable, and the combination of the set and the group can comprise a category when these satisfy category theory postulates [48,49]. That is because we believe that in future developments the discussion should embrace category theory as one of the important options.

To begin, according to our description for handling 'E's, it seems difficult to define inverse elements in a group theoretical way when there are deletions of 'E's from any place in a sequence because we cannot find sufficient numbers of 'E's in the target component of D_j.

Thus, we consider the 'morphism f' that transforms the sequence of DNA bases within set D as follows [48,49].

morphism $f : X \rightarrow X$, dom(f) = D_j, cod(f) = D_k. Object 'X' is the set of 'D_j's. There exists a morphism '1_X' such that '$1_X \bullet f = f = f \bullet 1_X$' for every 'morphism f', when '1_X' = $[n_1 | n_2 | n_3 | ... | n_i | ... | n_{N-1} | n_N | ...]$ ($\in$ group B). If supplemented, the 'morphisms f'

comprise 'group B' (see reference list in Figure 7). The group composition for 'f_1' and 'f_2' is denoted '$f_1 \bullet f_2$'.

As mentioned earlier, sequences of DNA consisting of bases 'C, A, T, and G' are transcribed into RNA consisting of 'C, A, U, and G'. This process can be regarded as the combination of two manipulation; I) transcription from the original DNA sequences (D_j) to those of its complement ($D_j^\dagger$), and II) alternation from 'C, A, T, and G' to 'C, A, U, and G' ($D_j^\dagger \rightarrow R_j^\dagger$) (both are illustrated in Figure 8).

Step I

The transformation from the original DNA bases 'D_j' into the complementary sequence '$D_j^\dagger$' (e.g., 'TCATEAGCTGA…' $\rightarrow$ 'AGTAETCGACT…') (for transcription to pre-messenger RNA (pre-mRNA) before splicing) can be performed via the manipulation (13–15, 17, 18, 20, 22, 23, 25a–c) in §3 and §4. '$D_j^\dagger$' can be obtained via the linear group (17, 18, 20), the rotational group (14, 15, 20) and also the wallpaper group (21–23, 25a–c, 26a–c). Thereby, morphism $\rho : X \rightarrow Y$, dom(ρ) = D_j, cod(ρ) = $D_j^\dagger$. Object 'Y' is the set of '$D_j^\dagger$'s (essentially equivalent to the set of 'D_j's).

There exist morphisms '1_X' and '1_Y' such that '$1_Y \bullet \rho = \rho = \rho \bullet 1_X$' for 'morphism ρ', where

$$\text{'}1_X = 1_Y = [n_1|n_2|n_3|...|n_i|...|n_{N-1}|n_N|...]\text{'}. \tag{47}$$

However, in practice, morphism ρ is one of the 'B_m's $\in$ group B (see Figures 7 and 8).

Morphism f : X→X. (f: $D_j \rightarrow D_k$). **Source object: X** (= set D), **target object: X** (= set D), dom(f) = D_j, cod(f) = D_k.   Practically, morphism f is one of the 'B_m's ($\in$ group B).

Morphism ρ : X→Y. (ρ: $D_j \rightarrow D_j^\dagger$). **Source object: X** (= set D), **target object: Y** (=set D), dom(ρ) = D_j, cod(ρ) = $D_j^\dagger$. Practically, morphism ρ is one of 'B_m's ($\in$ group B).

Morphism τ : Y→Z. (τ: $D_j^\dagger \rightarrow R_j^\dagger$). **Source object: Y** (= set D), **target object: Z** (= set R), dom(τ) = $D_j^\dagger$, cod(τ) = $R_j^\dagger$.

Morphism g = $\rho \bullet \tau$: X→Z. (g: $D_j \rightarrow R_j^\dagger$). **Source object: X** (= set D), **target object: Z** (= set R), dom(g) = D_j, cod(g) = $R_j^\dagger$.
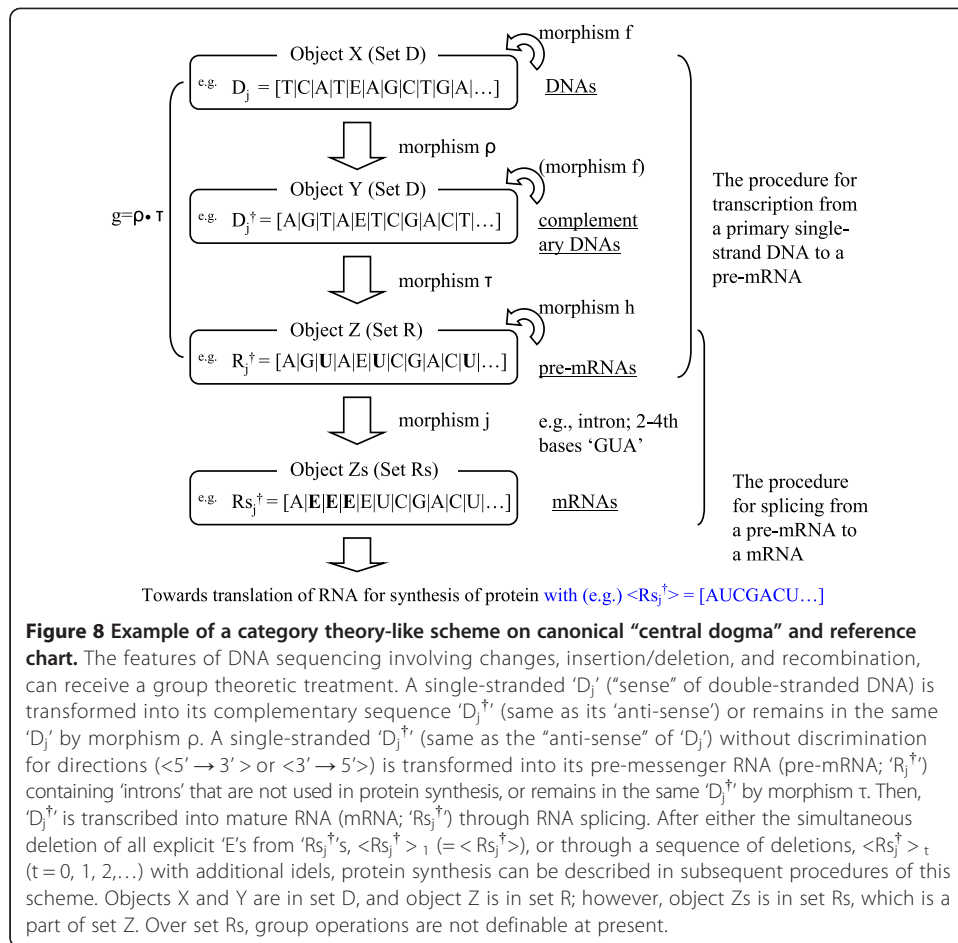
Morphism h: Z→Z. (h: '$R_j^\dagger \rightarrow R_j^{\dagger\prime}$' or '$R_j \rightarrow R_j$'). **Source object: Z** (= set R), **target object: Z** (= set R), dom(h) = $R_j^\dagger$, cod(h) = $R_j^\dagger$.   ($R_j^\dagger$, $R_j \in$ set R)

Like morphism f, morphism h is one of the 'B_m's ($\in$ group B).

Morphism j: Z→Zs. (j: '$R_j^\dagger \rightarrow Rs_j^{\dagger\prime}$' or '$R_j \rightarrow Rs_j$'). **Source object: Z** (= set R), **target object: Zs** (= set Rs), dom(j) = $R_j^\dagger$, cod(j) = $Rs_j^\dagger$.   ($Rs_j^\dagger$, $Rs_j \in$ set Rs)

Figure 7 Definition of category C. A simplistic definition of category C treating the traditional "central dogma" is presented. In practice, any selection of operations (morphisms) is permissible because composition within the category C using any of the three operations belonging to the 'linear group', 'rotational group' or 'wallpaper group' is considered possible. The morphisms f and h correspond to elements 'B_m' of group B, the only difference being the 'T's and 'U's. Actually, morphism j is regarded as part of morphism h ('$R_j^\dagger$'s and 'set R' are also substitutable for '$Rs_j^\dagger$'s and 'set Rs'); all morphisms except for 'τ' an 'g' satisfy the group postulates, and are treated as operations of group B.

Figure 8 Example of a category theory-like scheme on canonical "central dogma" and reference chart. The features of DNA sequencing involving changes, insertion/deletion, and recombination, can receive a group theoretic treatment. A single-stranded 'D$_j$' ("sense" of double-stranded DNA) is transformed into its complementary sequence 'D$_j^\dagger$' (same as its 'anti-sense') or remains in the same 'D$_j$' by morphism ρ. A single-stranded 'D$_j^\dagger$' (same as the "anti-sense" of 'D$_j$') without discrimination for directions (<5' → 3' > or <3' → 5'>) is transformed into its pre-messenger RNA (pre-mRNA; 'R$_j^\dagger$') containing 'introns' that are not used in protein synthesis, or remains in the same 'D$_j^\dagger$' by morphism τ. Then, 'D$_j^\dagger$' is transcribed into mature RNA (mRNA; 'Rs$_j^\dagger$') through RNA splicing. After either the simultaneous deletion of all explicit 'E's from 'Rs$_j^\dagger$'s, <Rs$_j^\dagger$> $_1$ (= < Rs$_j^\dagger$>), or through a sequence of deletions, <Rs$_j^\dagger$> $_t$ (t = 0, 1, 2,…) with additional idels, protein synthesis can be described in subsequent procedures of this scheme. Objects X and Y are in set D, and object Z is in set R; however, object Zs is in set Rs, which is a part of set Z. Over set Rs, group operations are not definable at present.

Step II

Next, we define manipulations that change the above 'D$_j^\dagger$' into 'R$_j^\dagger$' where all 'T's are converted into 'U's; e.g., (D$_j^\dagger$=) [A|G|**T**|A|E|**T**|C|G|A|C|**T**|…] → (R$_j^\dagger$=) [A|G|**U**|A|E|**U**|C|G|A|C|**U**|…]'.

This process can also be expressed in a similar way as transcription.

morphism τ : Y → Z, dom(τ) = D$_j^\dagger$, cod(ρ) = R$_j^\dagger$. Object 'Z' is the set of 'R$_j^\dagger$'s. There exist morphisms '1$_Y$' and '1$_Z$' such that '1$_Z$●τ = τ = τ●1$_Y$' for every 'morphism τ', where

$$'1_Y = 1_Z = [n_1|n_2|n_3|...|n_i|...|n_{N-1}|n_N|...]'. \tag{48}$$

(refer to Figures 7 and 8).

Evidently, morphism τ does not satisfy the group postulates because the source object 'Y' and target object 'Z' are different and a single set of operations cannot be defined at this stage.

Additionally, as for Steps I and II, the resultant process for morphisms ρ and τ can be expressed as:

$$\text{morphism } g = \rho●\tau : X→Z, \tag{49}$$

dom(g) = D$_j$, cod(g) = R$_j^\dagger$ (see Figures 7 and 8).

There exist morphisms '1$_X$' and '1$_Z$' such that '1$_Z$●g = g = g●1$_X$'.   (50)

The only difference between D$_j^\dagger$ and R$_j$ is the appearance 'T' and 'U' in the sequences.

Naturally, for RNA base sequences, similar treatments are possible in the single group B: morphism $h : Z \to Z$, $dom(h) = R_j^\dagger$, $cod(h) = R_k^\dagger$ (Figures 7 and 8).

There exists morphism '1_Z' such that '$1_Z \bullet h = h = h \bullet 1_Z$'. $\qquad$ (51)

Ordinarily, in prokaryotic cells, the DNA sequences are transcribed along their entire length. For eukaryotic cell, a splicing process is needed using nascent pre-messenger RNA (pre-mRNA) where introns of DNA bases are removed and exons are joined before producing a correct protein through translation, resulting in the mature messenger RNA (mRNA). Thus, the previous procedure was about the prokaryotic cell or the pre-translation of pre-mRNA in the eukaryotic cell. Therefore, to treat the products after this RNA splicing procedure in the eukaryotic cell, the following approach might be possible. The removal of introns can be regarded as changes from a certain series of bases to 'E's as follows.

If '**GUA**' is removed from 'A(**GUA**)EUCGACU...' to become 'A( )EUCGACU...', this procedure can be described as; '$R_j^\dagger \to Rs_j^\dagger$';

$$\text{'}R_j^\dagger = [A_1(\mathbf{G_2}|\mathbf{U_3}|\mathbf{A_4})E_5|U_6|C_7|G_8|A_9|C_{10}|U_{11}|\mathbf{E_{12}}|\mathbf{E_{13}}|\mathbf{E_{14}}|...]\text{'},\qquad(52)$$

$$\text{'}Rs_j^\dagger = [A_1(\mathbf{E_2}|\mathbf{E_3}|\mathbf{E_4})E_5|U_6|C_7|G_8|A_9|C_{10}|U_{11}|\mathbf{E_{12}}|\mathbf{E_{13}}|\mathbf{E_{14}}|...]\text{'}.\qquad(53)$$

The '$Rs_j^\dagger$' form a set $Rs = \{Rs_j^\dagger \ (j = 1,2,3,...)\}$ that is a part of set R (see Figures 7 and 8).

Hereon, we admit 'E's in the sequences of RNAs (as elements of set Rs) during the operations before morphism 'f' and after morphism 'j' to maintain theoretical consistency. Thus, if the result of a series of these maps is '$Rs_j^\dagger = A_1\mathbf{E_2E_3E_4}E_5U_6C_7G_8A_9C_{10}U_{11}\mathbf{E_{12}E_{13}E_{14}}...$'; then the actual RNA sequence should be interpreted as 'AUC...'. Specifically, an equivalent-sized substitution of some bases in pre-mRNA with 'E's can be written morphism $j : Z \to Zs$, $dom(j) = R_j^\dagger$, $cod(j) = Rs_j^\dagger$. There exists a morphism '1_{Zs}' such that '$1_{Zs} \bullet j = j = j \bullet 1_{Zs}$'.

'j' changes some series of bases from 'C, A, U, G, E' to an equivalent-sized series of 'E's within the partial operations of the group B. However, morphism 'j' fails the group axioms, as inverse might not be definable.

Finally, as in §1, we apply the simultaneous deletions of all explicit 'E's of mRNA other than the trailing 'E's, the state after these deletions being denoted with '< >'; for

$$\begin{aligned}\text{'}Rs_j^\dagger &= [A_1\mathbf{E_2E_3E_4}E_5U_6C_7G_8A_9C_{10}U_{11}\mathbf{E_{12}E_{13}E_{14}}...]\\ &= [A_1(\mathbf{E_2E_3E_4})E_5U_6C_7G_8A_9C_{10}U_{11}\mathbf{E_{12}E_{13}E_{14}}...]\text{'},\end{aligned}$$

the description '$<Rs_j^\dagger> = [A_1U_2C_3G_4A_5C_6U_7...]$' is specified without explicit non-trailing 'E's. In this regard, as in §1, if some indels (insertions/deletions) occur at certain bases of $<Rs_j^\dagger>$, as for '$<Rs_{j1}^\dagger> = [A_1U_2(\mathbf{E_3})G_4(\mathbf{E_5})C_6U_7...]$' (with the deletion of 'C_3' and 'A_5'), we state the result as '$<<Rs_{j1}^\dagger>> = [A_1U_2G_3C_4U_5...]$'. '$R_j^\dagger$'s include $<Rs_j^\dagger>$s and '$<<Rs_j^\dagger>>$s from the set R and both still satisfy the postulates of group B. This rule is a relative postulate and explicit 'E's are not absolutely forbidden in '$<Rs_j^\dagger>$'s or '$<<Rs_j^\dagger>>$'s, hence further indels of 'E's into '$<Rs_j^\dagger>$'s or '$<<Rs_j^\dagger>>$'s are not forbidden.

Also, omissions of explicit 'E's are considered as in '$\{<Rs_{j1}^\dagger>\} = [A_1U_2G_4C_6U_7...]$, where place numbers '3' and '5' are absent indicating implicitly their presence in the vector. (Note that all products belong to group B.) Similarly, t-tuples of '< >'s are denoted '$<<<<Rs_j^\dagger>>>>$ (t-tuple) $= <Rs_j^\dagger>_t$' representing multiple deletions of 'E's (t-times).

Combinations of symbols '{ }' and '< >' are also allowed when necessary, as for example {<{{<Rs_{j1}>}}>}, as long as the subscripted place numbers are adequately recognized/traced.

Nevertheless, the multiple use of '< >' to remove all 'E's in the vector '$Rs_j^\dagger$' should have a unique meaning with regard to protein synthesis. As a result, the subsequent reading/translation in line with 'codon'-like '[AUC|GAC|U...]'; '[A|UCG|ACU|...]' or '[AU|CGA|CU...]' leads in an ordinal way to a description of protein synthesis. Through the use of '{ }' and/or '< >', the concept 'E' may have benefits, although this may need to be intensely explored in future studies.

The procedure reversing transcription, found for example in retrovirus, is also describable if additional options are added to the scheme. However, these options are omitted at this stage to keep the model simple.

In summary, suppose we have a 'category C' with objects 'X', 'Y', 'Z', 'Zs' and morphisms 'f', 'ρ', 'τ', 'g', 'h' and 'j'. We affirm that these definitions satisfy the postulates of category. A list is given in Figure 7. Indeed, morphisms other than 'τ' and 'g' are simple group-theoretical products. One of the reasons we have introduced the concept 'category' is that the translation from single-strand DNAs to RNAs is difficult or impossible to systematize as a group structure. Therefore, if we identify the differences, we can treat all manipulations, except for 'τ' and 'g', based simply on group B.

The expression 'hom(X, X)' denotes all morphisms f: 'from X to X'. Likewise, 'hom(X, Y)' denotes all morphisms ρ: 'from X to Y'. In addition, 'hom(Y, Z)' denotes all morphisms τ: 'from Y to Z, and hom(X, Z) denotes all morphism g: from X to Z. Then, hom (Z, Z) denotes all morphisms h: 'from Z to Z'. Finally, hom(Y, Z) denotes all morphism h: from Y to Z. (Details are displayed in Appendix F)

As is explained in §3 and §4, the rotational group can be regarded as a specific bijection of the wallpaper group [2,44-47], so, we can describe this relationship naturally in a category theory-like way where two categories C_1 and C_2 are linked.

First, we consider two categories C_1 and C_2 with a 'functor F' from C_1 to C_2 written 'F: $C_1 \rightarrow C_2$'. For example, the pre-category C is denoted C_1 and the product of functor F on category C_1 is denoted C_2 [48,49]. Note that the only difference between C_1 and C_2 is assumed to be the nature of its expression; morphism $f_1 = B_1$ ($\in$ category C_1) is based on the wallpaper group in Figure 1 or 6; e.g.,

$$\text{'}B_1 = [r_1|l_2|u_3|...|n_i|...|r_{N-1}|d_N|...] = [...|x[a_i, \ b_i]_i...|] = [...|r^{ai}\bullet u^{bi}_i|...] \ (\in \text{ group } B_1)\text{'.} \tag{54}$$

Additionally, morphism $f_2 = B_2$ ($\in$ category C_2) is based on the rotational group over the Gaussian plane in Figure 5; e.g.,

$$\text{'}B_2 = [\omega_{1\ 1}|\omega_{4\ 2}|\omega_{2\ 3}|...|\omega_{0\ i}|...|\omega_{1\ N-1}|\omega_{3\ N}|...] = [...|(a_i + 2b_i)\omega_i|...] \ (\in \text{ group } B_2)\text{'.} \tag{55}$$

With regard to the identity morphisms, we have

$$\begin{aligned}\text{'}1_{X1} &= 1_{Y1} = 1_{Z1} = 1_{Zs1} = [n_1|n_2|n_3|...|n_i|...|n_{N-1}|n_N|...], 1_{X2} = 1_{Y2} = 1_{Z2} \\ &= 1_{Zs2} = [\omega_{0\ 1}|\omega_{0\ 2}|\omega_{0\ 3}|...|\omega_{0\ i}|...|\omega_{0\ N-1}|\omega_{0\ N}|...].\end{aligned} \tag{56}$$

Herein, we view 'functor F: $C_1 \rightarrow F(C_1)$ (= C_2)' in following way [48,49].
(Details are shown in Appendix G)

Note that a similar definition like the composition of C_1 based on the linear group and C_2 based on rotational group is possible, being linked with 'functor F'.

This is satisfied provided an adequate definition of 'Functor F' is given, and we presume that the morphisms described previously formulates a model that renders one of forms of the canonical "central dogma" proposed by Crick in 1958 [26].

In the transcription of RNA bases, the 'RNA splicing' process is well-known, whereby 'intron sequences' are excised, and 'exon sequences' are combined to condense effectual information for further interpretation in protein synthesis. However, as for further processing of the triplets of bases e.g., 'ACG' and 'AUG', a considerable number of models have been reported e.g., [12,18-21,33,34,50]. We refrain from pursuing this issue at present.

Results

We added an imaginary base 'E' to the set of actual DNA bases, and composed group Z_5 of basic translational operations on grid-points of a cruciform wallpaper pattern constructed of the five base letters. Moreover, using the same five letters, we integrated the wallpaper group as the combination of linear group over the horizontal line and the rotational group based on symmetries of a fivefold phasor diagram on the unit circle in the Gaussian plane. Additionally, changes in the sequences of the DNA bases are treated using set D, the set of all possible sequences of DNA bases that also contain 'E'. Also, 'D_j's are drawn as polygonal lines graphically. Moreover, by combining group Z_5, the operators that rearrange bases of DNA sequences constitute the group B. Using these results, simple changes of sequences, insertions/deletions, and recombination of DNA bases are also treatable via a synthesis of group-theoretical operations between sets D and group B. Together with this, all results obtained for DNA pertain to RNA by replacing T with U. Using these tools, category theory-like language is introduced to describe the canonical "central dogma" that is expected to integrate DNA-based processes, although the overall profile and range of applicability is unclear at this stage. Alternatively, by introducing the manipulations '{ }' and '< >', operations on states of 'E's in 'D_j's/'R_j's, whether explicit or not, can be performed in parallel with the conventional description for DNA/RNA sequences.

Discussion

The issue treated in this article is, roughly speaking, the combination of two ideas: one is the wallpaper pattern in the context of DNA sequencing '§1 – §7', and the other is the tentative development towards systematization of molecular/genetic biology in the style of some category-theory-like description '§8'. Essentially, the two are different topics although strongly connected. The former is an independent study on symmetry modeling of DNA sequences, whereas the latter can be re-expressed using different material as long as the basic elements can be treated within a category-theory-like model satisfying group theory-like postulates.

In this article, we considered a group/category theory-like treatment devising an expedient 'E', grid-point array, and group operations to move over the array. We discussed whether and how a more synthesized description can be constructed, using simplistic postulates of group. Next, we take the basics of category theory to describe processes, although this is only at a preliminary stage.

For an application of our ideas, we have chosen DNA sequencing from the perspective of not only coding sequences of DNA bases but also describing insertions/deletions of DNA sequences using a single operation that is an element of a group. The expeditious 'E' permits inserting and/or deleting sequences depending on the purpose. Specific notation was introduced so that vector-like DNA sequences and operations can be composed as a set and a group.

Ordinarily, a method to describe DNA sequences is often limited in scope by focusing on only one aspect such as recognizing each base sequentially (e.g., A, G, T, A, C…from 'AGTAC…') [13,33-35], where an operation like 'rotation', 'transition' or 'conversion' based on a certain solid is often used. Another focus of attention is the rules for interpretation of codons in synthesizing proteins from DNA sequences. The rules are defined to capture the specific activity from the viewpoint of group-like operations [18-22].

It also enables us to treat three manipulations as one type of operation in the group, with easily-imaginable graphic displays such as Figure 4, although it is only an accessorial tool at this stage. Increasing the degree of freedom by one and integrating changes of coding and sequence recombination might yield some polysemous utility.

When inserting/deleting sequences of bases into the main DNA sequence, even if the endpoints of the base series are identified precisely, it appears that manipulations via 'E's are not always necessary. Nonetheless, to determine the final order of the bases in these cases, we must track base changes from one to the next (including 'E', even when lost or deleted). If we use the 'E'-assisted manipulations for coding, we need only to examine the inclusions; the rest remains unchanged in order. Additionally, we assume that when any operation is performed, the position and number of 'E's should be fixed so that the order of any component of 'D_j' or 'B_j' is not changed, at least, during operations (e.g., (A.7)). The exception is specifically the insertion/deletion of 'E's such as in (29–34) and (36–45).

We briefly point out the notational benefits of the imaginary 'E's. These are three; 1) to adjust the sorts (number) of bases in DNAs and RNAs (from '4' to '5'), thus enabling group-theoretical composition over the (two-dimensional) plane; 2) to link the notational sequences of DNAs and mRNAs in a single format that can be used in a more compact database to record and analyze genetic information; and 3) to express sequences of DNAs/RNAs as a vector in three different ways: a) with explicit 'E's in the vector, b) with implicit 'E's in the vector, and c) with all 'E's omitted in the vector except the trailing 'E's. The last offers flexibility in storing world-wide genetic data in a single set. We suggest that exhaustiveness is one of the potentialities of the model adding versatility in addressing the possibilities of certain behaviors of DNA/RNA sequences. While that might be far from practical applications at this moment, a more rigorous methodology in the near future may yield a means.

Regarding style of the grid-point/cruciform/wallpaper pattern (Figure 1) in defining the group postulates, one of its advantages is that each base is surrounded by the four others. This symmetrical simplicity is absent in the linear group and the rotational group (Figures 1, 5 and 6), where the relative position of the five bases is fixed and thereby restrictive. Also, it might be crucial that the number '5' is key in enabling composition of the sort provided by the wallpaper group using the cruciform, and an identity element necessary to satisfy the group postulates. Being a prime, '5' will be convenient in further developments of the model exploiting algebraic structures such as rings or fields.

A similar synthesis might be possible between a modulo 7 additive rotational group based on a sevenfold phasor diagram with a space group depending upon six 'forward/backward', 'up/down' and 'left/right' directions. In practice, a space group is formed that consists of three orthogonal cruciforms comprising the six directions ($\pm x$, $\pm y$, $\pm z$) with seven elemental operations $\{m_u(\text{up}), m_d(\text{down}), m_r(\text{right}), m_l(\text{left}), m_f(\text{forward}), m_b(\text{backward})$ and $m_n(\text{no movement})\}$. These determine the operations of the group, which permute seven bases (prime number) or seven letter-like constituents. Analogous to the fivefold phasor diagram, we draw equispaced elements on the unit circle over the Gaussian plane; suppose '$\varphi = 2\pi/7$ (rad)', then the set $\{\varphi_1(= \varphi), \varphi_2(= 2\varphi), \varphi_3(= 3\varphi), \varphi_4(= 4\varphi), \varphi_5(= 5\varphi), \varphi_6(= 6\varphi), \varphi_0(= \varphi_7 = 0)\}$ parameterizes the rotational group [51], and both are, at least, in partial correspondence. We presume that, in extension, bringing together an n-dimensional space group (using the $2n + 1$ elements associated with the $\pm$ n-directions and E) and a rotational group based on the n-fold phasor diagram on the unit circle (with $2n + 1$ elements as points of the vertex of a polygon) might be possible. For this article, we have just focused on 'n = 2' in §1–§8.

Apart from the above, the model based on the wallpaper pattern might have a close relationship with 'cellular automata' [52]. Appropriate definitions of the wallpaper pattern for the five bases might find an expression between groups and cellular automata [53].

One consideration concerns whether a more integrated/synthesized style to describe biomolecular processes is possible using only simple, primitive defining rules, in particular, when describing genetic processes such as DNA transcription and RNA synthesis of proteins. Whereas the group postulates might be too restrictive to define molecular behavior, category postulates might enable such schemes to proceed because its postulates are weaker than those defining a group. If the interpretation of DNA by messenger RNA is definable within category theory, and protein synthesis is expressible within the same theory, there might be advantages in having the molecular system classified and treated in a reduced size in the database. At least, we conjecture that these ideas might be valid when clarifying impossible phenomena associated with changes of DNA sequences, resulting in reducing unnecessary, recrementitious efforts or roundabout paths that might encroach on researchers' limited time for investigation. That issue might be avoided if the impossibility of certain themes was known beforehand. From this standpoint, we believe that a mathematical systematization (in a general and unexceptional manner) is crucially important for future molecular/genetic biology.

The limitations of the present model should be noted. First, the wallpaper pattern drawn in Figure 1 is one example of various patterns. In general, the wallpaper groups have been classified into seventeen categories [2,44-47]. There could be other types of patterns like Figure 1 and groups upon which to compose this sort of model. For instance, if we exchange all 'A's for all 'C's, and all 'G's for all 'T's in the model presented in this article, an almost equivalent model '§1–§8' is constructed. Other arrangements might provide still unknown advantages that enable models like ours to be treated in a more rational manner. It remains unclear how to construct an optimal method to determine models yielding the wallpaper pattern of Figure 1 and the bijection given in Figure 3, and to develop the categories presented in Figures 7 and 8. The best positions of the five bases should be examined under a rigorous methodology.

Second, a Cartesian vector is defined as a combination of components on which operations are conducted independently. Indeed, we can perform operations on the

i-th component of D_j of set D using the (i + 1)-th component of B_j by adding an 'E' any place before the (i-1)-th component of D_j. This is because the components after the i-th of B_j shift to the right within the vector B_j. Therefore, in appearance, the components at different positions are essentially spectators (e.g., a base 'C_3' cannot change into either 'A_5' or G_5 by any B_j except via 'E'-assisted manipulations). In this case, after the insertion of two 'E's between the '**2**' and '**3**' components, '$C_{3+2}(= C_5)$' can become either 'A_5' or G_5' by acting appropriately on B_j at C_5. However, that might raise some confusion. For 'E'-assisted operations (such as 29–34, 38, 40), the results might change according to the place number of inserted/deleted 'E's that yields the mis-matches between the 'i-th' component of 'D_j' and that of 'B_j'. We believe further studies are warranted to find a descriptive format for the model.

Third, as for the graphical displays of 'D's in Figure 4, although the sequences of 'D_m's (m = 1, 2, 3) are in reality the same, the respective expressions are not always unique because the presence of the imaginary 'E's changes the shape of each sequence; e.g.,'$D_1 = [A_1C_2C_3(\ )G_4T_5\mathbf{E_6E_7E_8}...]$' and '$D_2 = [A_1C_2C_3(\mathbf{E_4E_5})G_6T_7\mathbf{E_8E_9E_{10}}...]$' are different over the wallpaper pattern despite being equivalent as real sequences. Although by use of electronic tools, these graphics might be of versatility for detection or identification of DNA sequences, these might produce other confusions in the present form. We hope that more appropriate devices would be performed in future study.

Fourth, DNA transcription to RNA and/or mRNA and translation of RNA and/or mRNA into proteins at the ribosomes are performed using a grammar rule based on a three-base set called a 'codon'. Codons have information to synthesize twenty types of proteins; for example, 'CAG' codes for 'glutamine'. As mentioned before, a number of approaches have been proposed exploit group-theoretic methods. These cover the rules for composition of triplet of bases 'XXX', the ways of reading codons, and models to compose geometric solids such as the tetrahedron and hexahedron, [12,18-21,33,34,50]. The rules for treating this aspect (transcription and translation of DNA bases' information) are not established in the present article. In addition, there are specific types of codon, such as 'TAA', 'TGA', and 'TAG', which are presently classified as 'stop' or 'halt' commands. Aside from this, there are various rules related to biogenetic activities such as DNA repair, alternative splicing, transposition, and translocation. These specific characterizations are lacking in our model, so, further improvements on this issue are desirable.

Fifth, the traditional symmetry model of DNA bases often is based on the chemical types 'purine/pyrimidine', 'amino/keto', and 'strong/weak hydrogen binding' using biomolecular characteristics, which often have advantages for their treatments where three-dimensional graphics aid the imagination, and `matricized' expressions are possible [29,35,36]. In our model, we merely use a rule for complementary pairing in §4 and §5. No restriction on couplings between 'C, A, T, G and E' is postulated in the present article. There might occur a number of combinations where non-realistic pairings of bases (e.g., 'A-G', 'C-C', and 'T-E') produce futilities and wastefulness in applications. We hope that future studies can solve this problem.

Sixth, there might be too many speculative conjectures with hypothetical situations those should be used to prove scientific facts using verified methods. Thus, a more

rigorous examination for a rational style with a more effective methodology is necessary.

Our model is far from a complete systematization. However, we believe that it is necessary that some principal breakthrough should be pursued if we intend to systemize a descriptive model, and that if appropriate definitions are devised, that might help to systemize biomolecular/genetic biology in a more optimized manner with greater sophistication to make a significant contribution to the field.

Conclusions

Within the large limitations of our methodology, it is considered that there is fertile ground where variants of the symmetry model for genetic coding based upon a specific wallpaper group are constructible. By integrating the linear group and rotational group over a specific wallpaper pattern, a more integrated formulation based on a group/category theory-like description is open to exploration in applications to a number of topics from molecular/genetic biology.

Appendix A

According to Figures 1, 3 and 6, the following relationships are confirmed straightforwardly between any bases and independently of the type of bases:

$$
\begin{aligned}
d \bullet d &= l \bullet u = u \bullet l = r \bullet n = n \bullet r = r, \\
r \bullet r &= d \bullet l = l \bullet d = u \bullet n = n \bullet u = u, \\
l \bullet l &= r \bullet u = u \bullet r = d \bullet n = n \bullet d = d, \\
u \bullet u &= r \bullet d = d \bullet r = l \bullet n = n \bullet l = l, \\
n \bullet n &= r \bullet l = l \bullet r = u \bullet d = d \bullet u = n
\end{aligned}
\tag{A.1}
$$

Here the symbol '↔' signifies 'bijection' and the meaning of '$x[-1, 0]$' is explained in §4. Hence, operators that are regarded to effect changes from one base to another can be re-expressed as illustrated in the following examples for various types of component operations:

$$
b_{[E \to C]} = b_{[C \to A]} = b_{[A \to T]} = b_{[T \to G]} = b_{[G \to E]} = r \ (\leftrightarrow \omega_1) = x[1, 0],
\tag{A.2}
$$

$$
b_{[E \to A]} = b_{[A \to G]} = b_{[G \to C]} = b_{[C \to T]} = b_{[T \to E]} = u \ (\leftrightarrow \omega_2) = x[0, 1],
\tag{A.3}
$$

$$
b_{[E \to T]} = b_{[T \to C]} = b_{[C \to G]} = b_{[G \to A]} = b_{[A \to E]} = d \ (\leftrightarrow \omega_3) = x[0, -1],
\tag{A.4}
$$

$$
b_{[E \to G]} = b_{[G \to T]} = b_{[T \to A]} = b_{[A \to C]} = b_{[C \to E]} = l \ (\leftrightarrow \omega_4) = x[-1, 0],
\tag{A.5}
$$

$$
b_{[E \to E]} = b_{[C \to C]} = b_{[A \to A]} = b_{[T \to T]} = b_{[G \to G]} = n \ (\leftrightarrow \omega_0 (= \textit{no rotation}) = \omega_5) = x[0, 0]
\tag{A.6}
$$

Appendix B

As for B,

1) Associativity: '$(B_j \bullet B_k) \bullet B_l = B_j \bullet (B_k \bullet B_l)$' holds for all positive integers j, k and l.

2) Identity: '$B_0 = [n_1 | n_2 | n_3 | \ldots | n_i | \ldots | n_{(n-1)} | n_n | n_{n+1} | n_{n+2} | n_{n+3} | \ldots]$' is an identity element that satisfies '$B_0 \bullet B_m = B_m \bullet B_0 = B_m$'. ($i = 1, 2, 3, \ldots$; '$n_i \ (=n)$' is an element of Z_5 (no movement of the point P))

3) Inverses: there exists a unique 'B_m^{-1}' that satisfies '$B_m^{-1} \bullet B_m = B_m \bullet B_m^{-1} = B_0$'.

Actually, the components of the inverse are the inverses of each individual component.

4) Commutativity: '$B_j \bullet B_k = B_k \bullet B_j$'.

5) Closure law: any '$B_j \bullet B_k$' belongs to the set B.

Appendix C

$$
\begin{aligned}
&D_2 \circ B_{(2 \to 3)} \bullet B_{(3 \to 4)} \\
&= [A_1|C_2|C_3(\mathbf{E_4}|\mathbf{E_5})G_6|T_7|A_N|\mathbf{E_8}|\mathbf{E_9}|...] \\
&\circ [n_1|n_2|n_3(\mathbf{u_4}|\mathbf{d_5})n_6|n_7|\mathbf{n_8}|\mathbf{n_9}|...] \bullet [l_1|u_2|\, d_3(\mathbf{r_4}|\mathbf{d_5})n_6|l_7|\mathbf{n_8}|\mathbf{n_9}|...], \\
&= [A_1 \circ n_1 \bullet l_1|C_2 \circ n_2 \bullet u_2|C_3 \circ n_3 \bullet d_3(\mathbf{E_5} \circ \mathbf{u_4} \bullet \mathbf{r_4}|\mathbf{E_5} \circ \mathbf{d_5} \bullet \mathbf{d_5})G_6 \circ n_6 \bullet n_6|T_7 \circ n_7 \bullet l_7|\mathbf{E_8} \circ \mathbf{n_8} \bullet n_8|\mathbf{E_9} \circ \mathbf{n_9} \bullet \mathbf{n_9}|...].
\end{aligned}
$$

$$(A.7)$$

Again, with reference to Figure 1 or Appendix B,

$$
\begin{aligned}
&= [A_1 \circ l_1|C_2 \circ u_2|C_3 \circ d_3(E_4 \circ d_4|E_5 \circ r_5)G_6 \circ n_6|T_7 \circ l_7|\mathbf{E_8} \circ \mathbf{n_8}|\mathbf{E_9} \circ \mathbf{n_9}|...], \\
&= [C_1|T_2|G_3(T_4|C_5)G_6|A_7|\mathbf{E_8}|\mathbf{E_9}|...] = D_4.
\end{aligned}
$$

$$(A.8)$$

Appendix D

Naturally, the series D_k is generated through the following sequence of operations:

$$
\begin{aligned}
B_{(j \to k)} &= [r^{4-1}{}_1|r^{3-2}{}_2|r^{1-0}{}_3|...|r^{4-1}{}_i|...|r^{1-3}{}_{N-1}|r^{3-2}{}_N|\mathbf{r^{0-0}{}_{N+1}}|\mathbf{r^{0-0}{}_{N+2}}|\mathbf{r^{0-0}{}_{N+3}}|...], \\
&= [r^3{}_1|r^1{}_2|r^1{}_3|...|r^3{}_i|...|r^{-2}{}_{N-1}|r^1{}_N|\mathbf{r^0{}_{N+1}}|\mathbf{r^0{}_{N+2}}|\mathbf{r^0{}_{N+3}}|...], \\
&= [r^3{}_1|r^1{}_2|r^1{}_3|...|r^3{}_i|...|r^{-2+5}{}_{N-1}|r^1{}_N|\mathbf{r^0{}_{N+1}}|\mathbf{r^0{}_{N+2}}|\mathbf{r^0{}_{N+3}}|...], \\
&= [r^3{}_1|r^1{}_2|r^1{}_3|...|r^3{}_i|...|r^3{}_{N-1}|r^1{}_N|\mathbf{r^0{}_{N+1}}|\mathbf{r^0{}_{N+2}}|\mathbf{r^0{}_{N+3}}|...].
\end{aligned}
$$

Then,

$$
\begin{aligned}
&D_j \circ B_{(j \to k)} \\
&= [C_1|A_2|E_3|...|C_i|...|T_{N-1}|A_N|\mathbf{E_{N+1}}|\mathbf{E_{N+2}}|\mathbf{E_{N+3}}|...] \circ [r^3{}_1|r^1{}_2|r^1{}_3|...|r^3{}_i|... \\
&...|r^3{}_{N-1}|r^1{}_N|\mathbf{r^0{}_{N+1}}|\mathbf{r^0{}_{N+2}}|\mathbf{r^0{}_{N+3}}|...], \\
&= [E \circ r^1{}_1|E \circ r^2{}_2|E \circ r^0{}_3|...|E \circ r^1{}_i|...|E \circ r^3{}_{N-1}|E \circ r^2{}_N|\mathbf{E \circ r^0{}_{N+1}}|\mathbf{E \circ r^0{}_{N+2}}|\mathbf{E \circ r^0{}_{N+3}}|...] \circ [r^3{}_1|r^1{}_2|r^1{}_3|...|r^3{}_i|... \\
&...|r^3{}_{N-1}|r^1{}_N|\mathbf{r^0{}_{N+1}}|\mathbf{r^0{}_{N+2}}|\mathbf{r^0{}_{N+3}}|...] = [E \circ r^{1+3}{}_1|E \circ r^{2+1}{}_2|E \circ r^{0+1}{}_3|...|E \circ r^{1+3}{}_i|... \\
&...|E \circ r^{3+3}{}_{N-1}|E \circ r^{2+1}{}_N|\mathbf{E \circ r^0{}_{N+1}}|\mathbf{E \circ r^0{}_{N+2}}|\mathbf{E \circ r^0{}_{N+3}}|...], = [E \circ r^{1+3}{}_1|E \circ r^{2+1}{}_2|E \circ r^{0+1}{}_3|...|E \circ r^{1+3}{}_i|... \\
&...|E \circ r^{3+3}{}_{N-1}|E \circ r^3{}_N|\mathbf{E \circ r^0{}_{N+1}}|\mathbf{E \circ r^0{}_{N+2}}|\mathbf{E \circ r^0{}_{N+3}}|...], \\
&= [E \circ r^4{}_1|E \circ r^3{}_2|E \circ r^1{}_3|...|E \circ r^4{}_i|...|E \circ r^6{}_{N-1}|E \circ r^3{}_N|\mathbf{E \circ r^0{}_{N+1}}|\mathbf{E \circ r^0{}_{N+2}}|\mathbf{E \circ r^0{}_{N+3}}|...], \\
&= [E \circ r^4{}_1|E \circ r^3{}_2|E \circ r^1{}_3|...|E \circ r^4{}_i|...|E \circ r^{6-5}{}_{N-1}|E \circ r^3{}_N|\mathbf{E \circ r^0{}_{N+1}}|\mathbf{E \circ r^0{}_{N+2}}|\mathbf{E \circ r^0{}_{N+3}}|...], \\
&= [E \circ r^4{}_1|E \circ r^3{}_2|E \circ r^1{}_3|...|E \circ r^4{}_i|...|E \circ r^1{}_{N-1}|E \circ r^3{}_N|\mathbf{E \circ r^0{}_{N+1}}|\mathbf{E \circ r^0{}_{N+2}}|\mathbf{E \circ r^0{}_{N+3}}|...], \\
&= [G_1|T_2|C_3|...|G_i|...|C_{N-1}|T_N|\mathbf{E_{N+1}}|\mathbf{E_{N+2}}|\mathbf{E_{N+3}}|...] = D_k.
\end{aligned}
$$

$$(A.9)$$

Appendix E

$D_j \circ B_{(j \to jt)}$
$= [E \circ x[0, 1]_1 | E \circ x[0, -1]_2 | E \circ x[0, 0]_3 | ... | E \circ x[1, 0]_i | ...$

$... | E \circ x[-1, 0]_{N-1} | E \circ x[0, 1]_N | \mathbf{E} \circ \mathbf{x[0, 0]_{N+1}} | \mathbf{E} \circ \mathbf{x[0, 0]_{N+2}} | \mathbf{E} \circ \mathbf{x[0, 0]_{N+3}} | ...]$

$\circ [x[0, -2]_1 | x[0, 2]_2 | x[0, 0]_3 | ... | x[-2, 0]_i | ... | x[2, 0]_{N-1} | x[0, -2]_N | \mathbf{x[0, 0]_{N+1}} | \mathbf{x[0, 0]_{N+2}} | \mathbf{x[0, 0]_{N+3}} | ...],$

$= [E \circ r^0 \bullet u^1{}_1 | E \circ r^0 \bullet u^{-1}{}_2 | E \circ r^0 \bullet u^0{}_3 | ... | E \circ r^1 \bullet u^0{}_i | ...$

$... | E \circ r^{-1} \bullet u^0{}_{N-1} | E \circ r^0 \bullet u^1{}_N | E \circ \mathbf{r^0} \bullet \mathbf{u^0}_{N+1} | E \circ \mathbf{r^0} \bullet \mathbf{u^0}_{N+2} | E \circ \mathbf{r^0} \bullet \mathbf{u^0}_{N+3} | ...],$

$\circ [r^0 \bullet u^{-2}{}_1 | r^0 \bullet u^2{}_2 | r^0 \bullet u^0{}_3 | ... | r^{-2} \bullet u^0{}_i | ... | r^2 \bullet u^0{}_{N-1} | r^0 \bullet u^{-2}{}_N | \mathbf{r^0} \bullet \mathbf{u^0}_{N+1} | \mathbf{r^0} \bullet \mathbf{u^0}_{N+2} | \mathbf{r^0} \bullet \mathbf{u^0}_{N+3} | ...],$

$= [E \circ r^0 \bullet u^1 \bullet r^0 \bullet u^{-2}{}_1 | E \circ r^0 \bullet u^{-1} \bullet r^0 \bullet u^2{}_2 | E \circ r^0 \bullet u^0 \bullet r^0 \bullet u^0{}_3 | ... | E \circ r^1 \bullet u^0 \bullet r^{-2} \bullet u^0{}_i | ...$

$| E \circ r^{-1} \bullet u^0 \bullet r^2 \bullet u^0{}_{N-1} | E \circ r^0 \bullet u^1 \bullet r^0 \bullet u^{-2}{}_N | E \circ \mathbf{r^0} \bullet \mathbf{u^0} \bullet \mathbf{r^0} \bullet \mathbf{u^0}_{N+1} | E \circ \mathbf{r^0} \bullet \mathbf{u^0} \bullet \mathbf{r^0} \bullet \mathbf{u^0}_{N+2} | E \circ \mathbf{r^0} \bullet \mathbf{u^0} \bullet \mathbf{r^0} \bullet \mathbf{u^0}_{N+3} | ...],$

$= [E \circ r^0 \bullet u^{-1}{}_1 | E \circ r^0 \bullet u^1{}_2 | E \circ r^0 \bullet u^0{}_3 | ... | E \circ r^{-1} \bullet u^0{}_i | ...$

$... | E \circ r^1 \bullet u^0{}_{N-1} | E \circ r^0 \bullet u^{-1}{}_N | E \circ \mathbf{r^0} \bullet \mathbf{u^0}_{N+1} | E \circ \mathbf{r^0} \bullet \mathbf{u^0}_{N+2} | E \circ \mathbf{r^0} \bullet \mathbf{u^0}_{N+3} | ...],$

$= [E \circ r^0 \bullet d^1{}_1 | E \circ r^0 \bullet u^1{}_2 | E \circ r^0 \bullet u^0{}_3 | ... | E \circ l^1 \bullet u^0{}_i | ...$

$... | E \circ r^1 \bullet u^0{}_{N-1} | E \circ r^0 \bullet d^1{}_N | E \circ \mathbf{r^0} \bullet \mathbf{u^0}_{N+1} | E \circ \mathbf{r^0} \bullet \mathbf{u^0}_{N+2} | E \circ \mathbf{r^0} \bullet \mathbf{u^0}_{N+3} | ...],$

$= [T_1 | A_2 | E_3 | ... | G_i | ... | C_{n-1} | T_n | \mathbf{E_{n+1}} | \mathbf{E_{n+2}} | \mathbf{E_{n+3}} | ...]$

$= D_j{}^\dagger.$

$$(A.10)$$

Appendix F
The axioms are:

I) A binary operation and closure law: the combination of two morphisms satisfies hom(X, X) × hom(X, Y) → hom(X, Y). Moreover, hom(X, Y) × hom(Y, Z) → mor (X, Z) and hom(Y, Z) × hom(Z, Zs) → mor (Y, Zs) both hold.

II) Associativity: If f: X → X, ρ: X → Y, τ: Y → Z, g: X → Z, h: Z → Z, and j: Z → Zs. Then, 'f●(ρ●τ) = (f●ρ)●τ', 'ρ●(τ●h) = (ρ●τ)●h', 'f●(g●h) = (f●g)●h', and 'τ●(h●j) = (τ●h)●j' hold.

III) Identity: there exist morphisms '1_X, 1_Y, 1_Z, 1_{Zs}' such that '1_X●f = f = f●1_X' and '1_Y●ρ = ρ = ρ●1_X', '1_Z●τ = τ = τ●1_Y', '1_Z●g = g = g●1_X', '1_Z●h = h = h●1_Z'. '1_{Zs}●j = j = j●1_Z'. In practice,

'$1_X = 1_Y = 1_Z = 1_{Zs} = [n_1 | n_2 | n_3 | ... | n_i | ... | n_{N-1} | n_N | ...]$' satisfies these conditions.

$$(A.11)$$

Appendix G
For Category C_1,

morphism $f_1 (= B_1 (\in \text{group } B_1))$: $X_1 \to X_1$,
morphism ρ_1: $X_1 \to Y_1$,
morphism τ_1: $Y_1 \to Z_1$,
morphism $g_1 (= \rho_1 \bullet \tau_1)$: $X_1 \to Z_1$,
morphism $h_1 (= B_1 (\in \text{group } B_1))$: $Z_1 \to Z_1$,
morphism j_1: $Z_1 \to Zs_1$.

$$(A.12)$$

Similarly for category C_2, for each object $F(X_1) = X_2$, $F(Y_1) = Y_2$, $F(Z_1) = Z_2$, $F(Zs_1) = Zs_2$ ($\in C_2$), the following relationships also hold:

$$
\begin{aligned}
&\text{morphism } F(f_1) \ (= f_2 = B_2(\in \text{group } B_2)) \colon F(X_1) \rightarrow F(X_1),\\
&\text{morphism } F(\rho_1) \ (= \rho_2) \colon F(X_1) \rightarrow F(Y_1),\\
&\text{morphism } F(\tau_1) \ (= \tau_2) \colon F(Y_1) \rightarrow F(Z_1),\\
&\text{morphism } F(g_1) \ (= g_2 = \rho_2 \bullet \tau_2) \colon F(X_1) \rightarrow F(Z_1),\\
&\text{morphism } F(h_1) \ (= h_2 = B_2(\in \text{group } B_2)) \colon F(Z_1) \rightarrow F(Z_1),\\
&\text{morphism } F(j_1) \ (= j_2) \colon F(Z_1) \rightarrow F(Zs_1).
\end{aligned}
\tag{A.13}
$$

Other than these, if relationships $F(f_1 \bullet \rho_1) = F(f_1) \bullet F(\rho_1)$, $F(\rho_1 \bullet \tau_1) = F(\rho_1) \bullet F(\tau_1)$, $F(\tau_1 \bullet h_1) = F(\tau_1) \bullet F(h_1)$, $F(f_1 \bullet g_1) = F(f_1) \bullet F(g_1)$, $F(g_1 \bullet h_1) = F(g_1) \bullet F(h_1)$, and $F(h_1 \bullet j_1) = F(h_1) \bullet F(j_1)$ are satisfied, the composition of C_1 and C_2 linked with 'functor F' is possible although the proof is omitted here.

Furthermore, the following postulates hold: for object X ($\in C_1$), '$F(1_X) = 1_{F(X)}$ ($\in C_2$)' is true, for object Y ($\in C_1$), '$F(1_Y) = 1_{F(Y)}$ ($\in C_2$)' and for object Z ($\in C_1$), '$F(1_Z) = 1_{F(Z)}$ ($\in C_2$)', is true under the condition:

$$
'F(1_X) = 1_{F(X)} = F(1_Y) = 1_{F(Y)} = F(1_Z) = 1_{F(Z)} = [\omega_{0\ 1}|\omega_{0\ 2}|\omega_{0\ 3}|...|\omega_{0\ i}|...|\omega_{0\ N-1}|\omega_{0\ N}|...]'.
\tag{A.14}
$$

Author details
[1]Department of Psychiatry, Tokyo Women's Medical University, Tokyo, Japan. [2]Depression Prevention Medical Center, Inariyama Takeda Hospital, Kyoto, Japan.

References
1. Rosen J: *Symmetry rules: How science and nature are founded on symmetry.* 1st edition. New York: Springer-Verlag; 2008.
2. Armstrong MA: *Groups and symmetry, undergraduate texts in mathematics.* New York: Springer-Verlag; 1988.
3. Judson TW: *Abstract Algebra: Theory and Applications.* Virginia: PWS Publishing Company; 1997.
4. Hungerford TW: *Abstract Algebra, An Introduction.* 2nd edition. Philadelphia: Saunders College Publishing; 1997.
5. Tung WK: *Group theory in physics.* Singapore: World Scientific Pub. Co. Inc.; 1985.
6. Unger AA: **The abstract Lorentz transformation group.** *Am J Phy* 1992, **60**:815–828.
7. Sexl RU, Ulbantke HU: *Relativity, Groups, Particles: Special Relativity and Relativistic Symmetry in Field and Particle Physics.* Wien: Springer; 2001.
8. Hamermesh M: *Group theory and its application to physical problems.* New York: Dover Publications, Inc.; 2012.
9. Ladd M: *Symmetry and group theory in chemistry.* Cambridge: Woodhead Publishing Limited; 1998.
10. Derome J-R: **Biological Similarity and Group Theory.** *J Theor Biol* 1977, **65**:369–378.
11. Wang J: **A complete symmetrical group DNA sequences and symmetry of poly-codon sequences (I).** *J Biomathematics* 2001, **16**:129–136.
12. Wang J: **A complete symmetrical group DNA sequences and symmetry of poly-codon sequences (II).** *J Biomathematics* 2001, **16**:257–265.
13. Chirikjian GS: **Group theory and biomolecular conformation: I Mathematical and computational models.** *J Phys Condens Matter* 2010, **22**:323103.
14. Chirikjian GS: **Mathematical aspects of molecular replacement. I. Algebraic properties of motion spaces.** *Acta Crystallogr A* 2011, **67**:435–436.
15. Fischer M, Klaere S, Nguyen MAT, von Haeseler A: **On the group theoretical background of assigning stepwise mutations onto phylogenies.** *Algorithms Mol Bio* 2012, **7**:36.
16. Bashford JD, Tsohantjis I, Jarvis PD: **Codon and nucleotide assignments in a supersymmetric model of the genetic code.** *Phys Lett A* 1997, **233**:288–481.

17. Bashford JD, Tsohantjis I, Jarvis PD: **A supersymmetric model for the evolution of genetic code.** *Proc Natl Acad Sci U S A* 1998, **95**:987–992.
18. Sánchez R, Morgado E, Grau R: **Gene algebra from a genetic code algebraic structure.** *J Math Biol* 2005, **51**:431–457.
19. Sánchez R, Grau R: **Vector space of the extended base-triplets over the Galois field of five DNA bases alphabet.** *World Acad Sci Eng Technol, Int J Comp, Inf Sci Eng* 2007, **1**:5.
20. Sánchez R, Grau R: **An algebraic hypothesis about the primeval genetic code architecture.** *Math Biosci* 2009, **221**:60–76.
21. Sánchez R, Grau R: **A novel Lie algebra of the genetic code over the Galois field of four DNA bases.** *Math Biosci* 2006, **202**:156–174.
22. Rietman EA, Karp RL, Tuszynski JA: **Review and application of group theory to molecular systems biology.** *Theor Biol and Med Modell* 2011, **8**:21.
23. Korn F: *Elementary Structures reconsidered: Lévi-Strauss on kinship.* London: Routledge; 2004.
24. Crick FHC: **Codon-anticodon pairing: The wobble hypothesis.** *J Mol Biol* 1966, **19**:548–555.
25. Crick FHC: **The origin of the genetic code.** *J Mol Biol* 1968, **38**:367–379.
26. Crick FHC: **On protein synthesis.** *In Symp Soc Exp Biol* 1958, **12**:138–163.
27. Bashford JD, Jarvis PD: **The genetic code as a periodic table.** *Biosystems* 2000, **57**:147–161.
28. Bíró T, Czirók A, Vicsek T, Major B: **A application of vector space techniques to DNA.** *Fractals* 1998, **6**:205–210.
29. Beland P, Allen TF: **The origin and evolution of the genetic code.** *J Theor Biol* 1994, **170**:359–365.
30. Epstein CJ: **Role of the amino-acid "code" and of selection for conformation in the evolution of proteins.** *Nature* 1966, **210**:25–28.
31. Jukes TH, Osawa S: **Evolutionary changes in the genetic code.** *Comp Biochem Physiol* 1993, B **106**:489–494.
32. Jukes TH: **Evolution of the amino acid code: Inferences from mitochondrial codes.** *J Mol Evol* 1983, **19**:219–225.
33. Pickover CA: **DNA and protein tetragrams.** *J Mol Graphics* 1992, **10**:2–6.
34. Trainor LEH, Rowe GW, Szabo VL: **A tetrahedral representation of poly-codon sequences and a possible origin of codon degeneracy.** *J Theor Biol* 1984, **108**:459–468.
35. Zhang R, Zhang CT: **Z curves, an intuitive tool for visualizing and analyzing the DNA sequences.** *J Biomol Str Dyn* 1994, **11**:767–782.
36. Zhang CT: **A symmetrical theory of DNA sequences and its applications.** *J Theor Biol* 1997, **187**:297–306.
37. Duplij D, Duplij S: **DNA sequence representation by trianders and determinative degree of nucleotides.** *J Zhejiang Univ (Sci)* 2005, **6B**:743–755.
38. Rushdi A, Tuqan J, Strohmer T: **Map-invariant spectral analysis for the identification of DNA periodicities.** *EURASIP J Bioinform Syst Biol* 2012, **2012**:16.
39. Zupan J, Randić M: **Algorithm for coding DNA sequences into "spectrum-like" and "Zigzag" representations.** *J Chem Inf Model* 2005, **45**:309–313.
40. Jafarzadeh N, Iranmanesh A: **A novel graphical and numerical representation for analyzing DNA sequences based on codons.** *MATCH Commun Math Comput Chem* 2012, **68**:611–620.
41. He P, Wang J: **Numerical characterization of DNA primary sequence.** *Internet Elec J Mol Des* 2002, **1**:668–674.
42. Nandy A, Harle M, Basak SC: **Mathematical descriptors of DNA sequences: Development and applications.** *ARKIVOC* 2006, **ix:**211–238.
43. Sawamura J, Morishita S, Ishigooka J: **A group-theoretical notation for disease states: an example using the psychiatric rating scale.** *Theor Biol Med Model* 2012, **9**:28.
44. Martin GE: *Transformation Geometry.* New York: Springer-Verlag; 1983.
45. Lockwood EH, Macmillan RH: *Geometric symmetry.* Cambridge (England): Cambridge University press; 1978.
46. Liu Y, Collins RT: **Skewed symmetry groups.** In *Proc. IEEE Conf. Computer Vision and Pattern Recognition.* ; 2001:872–879.
47. Liu Y: **A computational model for periodic pattern perception based on frieze and wallpaper groups.** *IEEE Trans PAMI* 2004, **26**(3):354–371.
48. Awodey S: *Category Theory (Oxford Logic Guides).* 2nd edition. New York: Oxford University Press, Inc.; 2010.
49. Mac Lane S: *Categories for the working mathematician.* 2nd edition. New York: Springer-Verlag; 1998.
50. D'Onofrio D, Abel D, Johnson DE: **Dichotomy in the definition of prescriptive information suggests both prescribed data and prescribed algorithms: biosemiotics applications in genomic systems.** *Theoret Biol Med Modell* 2012, **9**:8.
51. Chung KW, Chan HSY, Wang BN: **Tessellations with symmetries of the wallpaper groups and the modular group in the hyperbolic 3-space form dynamics.** *Comput Graph* 2001, **25**:333–341.
52. Butler JT: **A note on cellular automata simulations.** *Inf Control* 1974, **26**:286–295.
53. Ceccherini-Silberstein TG, Machi A, Scarabotti F: **Amenable groups and cellular automata.** *Ann Inst Fourier (Genoble)* 1999, **49**(issue 2)**:**673–685.